

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
- TEXT CUT OFF AT TOP, BOTTOM OR SIDES
- FADED TEXT
- ILLEGIBLE TEXT
- SKEWED/SLANTED IMAGES
- COLORED PHOTOS
- BLACK OR VERY BLACK AND WHITE DARK PHOTOS
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 : Search time 13.078 Seconds
(without alignments)
1283.801 Million cell updates/sec

Title: US-10-007-267-2
Perfect score: 598
Sequence: 1 LQAVAVFKQLPEAAALAAAN.....LNRRLAQMNVAADIALIGE 126

Scoring table:
Gapop 10.0, Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

1: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB	ID	Description
1	598	100.0	126	18	AAW06577	Lipo-oligosacchari
2	164	27.4	672	23	ABBS4412	Lactococcus lactis
3	163	27.3	989	22	AAU33401	Enterococcus faeca
4	161	26.9	678	18	AAW33905	Streptococcus pneu
5	161	26.9	678	19	AAW37953	Amino acid sequenc
6	154	25.8	679	23	ABP26830	Streptococcus poly
7	135	22.6	100	23	ABP26829	Streptococcus poly
8	135	22.6	679	23	ABP28806	Streptococcus poly
9	132.5	22.2	80	23	ABP07415	Human ORFX protein
10	119	19.9	1067	23	ABBS2588	Herbicidally activ

11	103.5	17.3	688	23	ABBA9111	Listeria monocytog
12	88	14.7	1503	20	AAV27142	Human slit-2 matur
13	88	14.7	1503	20	AAW6701	Slit-like protein
14	88	14.7	1525	20	AAV17499	Human Slit-1 prote
15	88	14.7	1529	20	AAV27145	Human slit-2 prote
16	88	14.7	1529	20	AAW6702	Full length slit-1
17	88	14.7	1529	22	AAW61239	Human Slit-1 prote
18	87	14.5	1529	21	AAV76117	Rat Slit homologue
19	87	14.5	1529	22	AAW56056	Skin cell protein,
20	87	14.5	1529	23	ABW72356	Rat protein isolat
21	85	14.2	1530	21	AAW28151	Slit protein. Xen
22	83	13.9	599	22	ABG24227	Novel human diagno
23	82.5	13.8	484	22	AAW78985	Human protein SEO
24	82.5	13.8	533	22	AAW79969	Human protein SEO
25	82.5	13.8	620	21	AAW51261	Arabidopsis thalia
26	82.5	13.8	638	21	AAW51260	Arabidopsis thalia
27	82.5	13.8	754	21	AAW51259	Arabidopsis thalia
28	82.5	13.8	1374	22	AAW69070	Human male enhance
29	82	13.7	232	21	AAV74904	Neisseria meningit
30	82	13.7	232	21	AAV74907	Neisseria meningit
31	82	13.7	815	23	ABP27680	Streptococcus poly
32	82	13.7	815	23	ABP29822	Streptococcus poly
33	80.5	13.5	232	21	AAV74906	Neisseria meningit
34	80.5	13.5	1224	22	ABW62191	Drosophila melanog
35	80	13.4	231	21	AAV74903	Neisseria meningit
36	79.5	13.3	594	21	AAV44280	Murine A20 Binding
37	79.5	13.3	647	21	AAV44279	Murine A20 Binding
38	79.5	13.3	647	21	AAV44285	Mutant A20 Binding
39	79.5	13.3	647	21	AAV44286	Mutant A20 Binding
40	79.5	13.3	647	21	AAV44287	Mutant A20 Binding
41	79.5	13.3	647	21	AAV44288	Mutant A20 Binding
42	79	13.2	888	23	AAW83013	Human homologue of
43	78.5	13.1	255	18	AAW20772	H. pylori cytoplas
44	78.5	13.1	1030	21	AAW38915	Arabidopsis thalia
45	78.5	13.1	1133	21	AAW38914	Arabidopsis thalia

ALIGNMENTS

RESULT 1	AAW06577	standard; Protein: 126 AA.
XX	AAW06577:	
AC	21-MAR-1997	(first entry)
XX	Lipo-oligosaccharide gene-encoded protein.	
DE	Polyglycosyltransferase; N-acetylglucosaminyl transferase:	
XX	N-acetylglucosaminyl transferase; lipo-oligosaccharide.	
KW	Neisseria gonorrhoeae ATCC 33084.	
XX	W09640971-A1.	
OS	19-DEC-1996.	
XX	03-JUN-1996;	96W0-US08323.
XX	07-JUN-1995;	95US-0478140.
PR	(NEOS-) NEOFSE TECHNOLOGIES INC.	
PA	Buczala SL, Johnson KF, Roth S;	
XX	WPI: 1997-052351/05.	
PI	N-PSDB: AAT49230.	
XX	Transfer of at least 2 saccharide units using	
DR	poly:glycosyl:transferase - isolated from N. gonorrhoeae, catalyses	
XX	the addition of both GlcNAc and GalNAc di:saccharide(s) units to a	
PT		

PT single galactose moiety
XX
PS Disclosure; Fig 2A; 38pp; English.
XX
CC A lipooligosaccharide-encoding gene region (AAT49230) of *Neisseria*
CC gonorrhoeae ATCC 33084 includes coding sequences for 5 proteins
CC (AAW06576-80), one of which (AAW06576) is a polyglycosyltransferase
CC that catalyses the addition of GlcNAc and GalNAc disaccharides to
CC a galactose moiety. The function of the other 4 proteins is not
CC stated in the specification.
XX
SQ Sequence 126 AA;
Query Match 100.0%; Score 598; DB 18; Length 126;
Best Local Similarity 100.0%; Pred. No. 1.3e-55;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 LQAAVAFKQLPDEAALAAANKRVONLKKADALGEVNESLLOODEERKALYAAAGLOPK 60
DB 1 LQAAVAFKQLPDEAALAAANKRVONLKKADALGEVNESLLOODEERKALYAAAGLOPK 60
QY 61 IAAVABEGNFTALSELASVSRPOVDAPFDGYVMVAEDAAVAKONRLNLRLAEQNNAYAD 120
DB 61 IAAVABEGNFTALSELASVSRPOVDAPFDGYVMVAEDAAVAKONRLNLRLAEQNNAYAD 120
QY 121 IALLGE 126
DB 121 IALLGE 126
Db 121 IALLGE 126
RESULT 2
ABB54412
ID ABB54412 standard; Protein: 672 AA.
XX
AC ABB54412;
DT 16-MAY-2002 (first entry)
XX
DE Lactococcus lactis protein glusB.
XX
KW Biosynthesis; biodegradation; lactic bacterium; yogurt; cheese.
XX
OS Lactococcus lactis IL1403.
XX
PN FR2807446-A1.
XX
PD 12-OCT-2001.
XX
PF 11-APR-2000; 2000FR-0004630.
XX
PR 11-APR-2000; 2000FR-0004630.
XX
PA (INRG) INRA INST NMT RECH AGRONOMIQUE.
PI Bolotine A, Sorokine A, Renault P, Ehrlich SD;
PI WPI; 2002-043418/06.
DR
XX
PT New nucleotide sequence useful in the identification or *Lactococcus*
PT lactis and related species -
XX
XX
PS Claim 6; SEQ ID No 1114; 2504pp; French.
XX
CC The present invention is related to a *Lactococcus lactis* nucleotide
CC sequence (ABA90521) and related proteins (ABB53300-ABB55621). The
CC nucleic acid sequence is useful in the detection and/or amplification of
CC nucleic acid sequence, particularly to identify *Lactococcus lactis* or
CC related species. The proteins of the invention are useful for the
CC biosynthesis or biodegradation of a composition of interest. The
CC invention helps research in lactic bacteria, particularly useful in the
CC production of yogurt and cheese.
CC Note: The sequence data for this patent is based on equivalent patent
CC WO20017734 (published 18-OCT-2001) which is available in electronic

CC format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.
XX
SQ Sequence 672 AA;
Query Match 27.4%; Score 164; DB 23; Length 672;
Best Local Similarity 38.0%; Pred. No. 9.6e-09;
Matches 38; Conservative 21; Mismatches 37; Indels 4; Gaps 2;
QY 25 NLKKADALGEVNESLLOODEERKALYAAAGLOPKIAAAVABEGNFTALSELASVSRPOV 84
DB 575 NLVAKA-KDIKEINPTLFEEDAEALYNNVISLONQTYMPGEERFALYHSLA---PAI 630
QY 85 DAFPDGYVMVAEDAAVAKONRLNLRLAEQNNAYADALL 124
DB 631 ETFESVYMAEDLSVRDNRITALLSEVVALTSVMADESLI 670
RESULT 3
ID ABB54412 standard; Protein: 989 AA.
XX
AC ABB54412;
DT 14-FEB-2002 (first entry)
XX
DE Enterococcus faecalis cellular proliferation protein #37.
XX
KW Antisense; prokaryotic cellular proliferation protein;
KW antibiotic; antibacterial; drug design.
XX
OS Enterococcus faecalis.
XX
PN WO200170955-A2.
XX
PD 27-SEP-2001.
XX
PF 21-MAR-2001; 2001WO-US09180.
XX
PR 21-MAR-2000; 2000US-191078P.
XX
PR 23-MAY-2000; 2000US-206848P.
XX
PR 26-MAY-2000; 2000US-207727P.
XX
PR 23-OCT-2000; 2000US-242578P.
XX
PR 27-NOV-2000; 2000US-253625P.
XX
PR 22-DEC-2000; 2000US-257931P.
XX
PR 16-FEB-2001; 2001US-269308P.
XX
PA (ELIT-) ELITRA PHARM INC.
XX
PI Haselbeck R, Ohlsen KL, Zyskind JW, Wall D, Trawick JD, Carr GJ;
PI Yamamoto RT, Xu HH;
XX
DR WPI; 2001-611495/70.
XX
DR N-PSDB; AAS51260.
XX
PT New polynucleotides for the identification and development of
PT antibiotics, comprise sequences of antisense nucleic acids -
XX
XX
PS Example 3; Seq ID No 4897; 511pp; English.
XX
CC The invention relates to antisense inhibitors of genes essential to
CC prokaryotic cellular proliferation, their use in identifying the
CC genes, their use in the discovery of novel antibiotics, the essential
CC genes themselves and the encoded proteins. The prokaryotes used are
CC *Escherichia coli*, *Staphylococcus aureus*, *Salmonella typhi*, *Klebsiella*
CC *pneumoniae*, *Pseudomonas aeruginosa* and *Enterococcus faecalis*. The
CC invention is also useful for the identification of potential new targets
CC for antibiotic development. The antisense nucleic acids can also be used
CC to identify proteins used in proliferation, to express these proteins,
CC and to obtain antibodies capable of binding to the expressed proteins.
CC The proteins can be used to screen compounds in rational drug discovery
CC programmes. The antisense nucleic acid sequence is also useful to screen
CC for homologous nucleic acids which are required for cell proliferation in
CC a wide variety of organisms. The present sequence represents an

CC antagonists with potential antibacterial activity. The
CC polynucleotides are also useful for DNA vaccination

```

XX Sequence      678 AA;
SQ
Query Match      26.9%; Score 161; DB 19; Length 678;
Best Local Similarity 35.5%; Pred. No. 2e-08;
Matches 43; Conservative 21; Mismatches 37; Indels 20; Gaps 3;

OY 6 VFQQLPEAAALAAANK-----RVQNLKKADALGEVNSLLQODEKALYAA 53
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 553 VADMLEAASALVEYSKEDFKPSVESLSRAFNLEKAE-GVATYDSALFENDQEKALAE 611
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 54 AQLQPKIAAAVAEGNFRFALSIAVKPQVDAPFDGVVMAEDPAVKQNRLLNRLAE 113
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 612 VETL-----VLGGPASQQLKQLFALSPYIDAFEFNTVMVAEDQAVRQNRLLATLSQLTK 664
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 114 Q 114
DB 665 K 665

```

RESULT 6

ABP26830
ID ABP26830 standard; Protein; 679 AA.

XX ABP26830;

XX 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 2836.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus pyogenes.

XX MO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

XX Tettelein H;

XX WPI: 2002-352536/38.

XX N-PSDB; ABN67461.

XX Claim 1; Page 3435; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in the specification. The proteins have antibacterial and antiinflammatory activity. (I), nucleic acids encoding (I), ABN65044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by Streptococcus bacteria, particularly S. agalactiae and S. pyogenes. Nucleic acids encoding (I) are used to detect Streptococcus in a biological sample. (I) is used to determine whether a compound binds to (I). A composition comprising (I) or a nucleic acid encoding (I), may be used as a vaccine or diagnostic composition. The disease caused by

CC Streptococcus that is prevented or treated may be meningitis. Nucleic acid encoding (I) may be used to recombinantly produce (I) and may be used in gene therapy. Antibodies to (I) are used for affinity chromatography, immunoassays, and distinguishing/identifying Streptococcus proteins.

XX Sequence 679 AA;

Query Match 25.8%; Score 154; DB 23; Length 679;
Best Local Similarity 33.3%; Pred. No. 1.1e-07;
Matches 41; Conservative 25; Mismatches 37; Indels 20; Gaps 4;

```

OY 8 KQLEBAALAAAN-----KRVNLLKKADALGEVNSLLQODEKALYAAQ 55
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 556 EMLAAEALVYASHTENKPRVESLSRAFNLEKADASV-QVDSLEENEGENTLFAIQ 614
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 56 GLQPKIAAAVAEGNFRFALSIAVKPQVDAPFDGVVMAEDPAVKQNRLLNRLAEQ 115
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 615 GL--TLGSAAQ-----QLQGVFALSPYINDFPNTVMAGDQALKNNRLATLSDLVSKA 667
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
OY 116 NAV 118
DB 668 KTI 670

```

RESULT 7

ABP26829
ID ABP26829 standard; Protein; 100 AA.

XX ABP26829;

XX 02-JUL-2002 (first entry)

DE Streptococcus polypeptide SEQ ID NO 2834.

XX Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;

KW group A streptococcus; Streptococcus pyogenes; antibacterial;

KW antiinflammatory; infection; vaccine; meningitis; gene therapy.

OS Streptococcus agalactiae.

XX MO200234771-A2.

XX 02-MAY-2002.

XX 29-OCT-2001; 2001WO-GB04789.

XX 27-OCT-2000; 2000GB-0026333.

XX 24-NOV-2000; 2000GB-0028727.

XX 07-MAR-2001; 2001GB-0005640.

XX (CHIR-) CHIRON SPA.

XX (GENO-) INST GENOMIC RES.

XX Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;

XX Tettelein H;

XX WPI: 2002-352536/38.

XX N-PSDB; ABN67460.

XX Claim 1; Page 3435; 4525pp; English.

The invention relates to a protein (ABP25413-ABP30895) from group B streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in the specification. The proteins have antibacterial and antiinflammatory activity. (I), nucleic acids encoding (I), ABN65044-ABN71526 and antibodies that bind (I) are used in the manufacture of medicaments for the treatment or prevention of infection or disease caused by

CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 XX Sequence 100 AA:
 SQ
 Query Match 22.6%; Score 135; DB 23; Length 100;
 Best Local Similarity 35.0%; Pred. No. 9.9e-07;
 Matches 36; Conservative 18; Mismatches 41; Indels 8; Gaps 2;
 OY 22 RVQNLKKADALCEVNESLIQODEKALYAAAGLOPKIAAAYAEQNFRTALSELASVK 81
 DB 3 RAFNLAEKVTHSV-LVDSLSFENNOEKALYQAISLE-----LTEDMHNDLKLFLALS 54
 OY 82 PQVDAFEDGVVMAEDAAYKONRLNLRLAEQNNAYADIALL 124
 DB 55 PIIDFEDNTWMTDDEKMKONRLAILNSIVAKARTVAAFNML 97
 RESULT 8
 ABP28806
 ID ABP28806 standard; Protein: 679 AA.
 XX
 AC ABP28806;
 XX
 DT 02-JUL-2002 (first entry)
 XX
 DE Streptococcus polypeptide SEQ ID NO 6788.
 XX
 KM Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
 KM group A streptococcus; Streptococcus pyogenes; antibacterial;
 KM antiinflammatory; infection; vaccine; meningitis; gene therapy.
 XX
 OS Streptococcus agalactiae.
 XX
 PN WO200234771-A2.
 XX
 PD 02-MAY-2002.
 XX
 PF 29-OCT-2001; 2001WO-GB04789.
 XX
 PR 27-OCT-2000; 2000GB-0026333.
 PR 24-NOV-2000; 2000GB-0028727.
 PR 07-MAR-2001; 2001GB-0005640.
 XX
 PA (CHIR-) CHIRON SPA.
 PA (GENO-) INST GENOMIC RES.
 PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
 PI Tetelin H;
 XX
 DR WPI; 2002-352536/38.
 DR N-PSDB; ABN69437.
 XX
 XX New Streptococcus protein for the treatment or prevention of infection
 PT or disease caused by Streptococcus bacteria, such as meningitis, and
 PT for detecting a compound that binds to the protein -
 XX
 Claim 1; Page 3841; 4525pp; English.
 XX
 CC The invention relates to a protein (ABP25413-ABP30895) from group B
 CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
 CC (Streptococcus pyogenes), comprising one of 5483 sequences (SI), given in
 CC the specification. The proteins have antibacterial and antiinflammatory
 CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
 CC antibodies that bind (I) are used in the manufacture of medicaments for

CC the treatment or prevention of infection or disease caused by
 CC Streptococcus bacteria, particularly *S. agalactiae* and *S. pyogenes*.
 CC Nucleic acids encoding (I) are used to detect Streptococcus in a
 CC biological sample. (I) is used to determine whether a compound binds to
 CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
 CC used as a vaccine or diagnostic composition. The disease caused by
 CC Streptococcus that is prevented or treated may be meningitis. Nucleic
 CC acid encoding (I) may be used to recombinantly produce (I) and may be
 CC used in gene therapy. Antibodies to (I) are used for affinity
 CC chromatography, immunoassays, and distinguishing/identifying
 CC Streptococcus proteins.
 XX
 XX Sequence 679 AA:
 SQ
 Query Match 22.6%; Score 135; DB 23; Length 679;
 Best Local Similarity 35.0%; Pred. No. 1.1e-05;
 Matches 36; Conservative 18; Mismatches 41; Indels 8; Gaps 2;
 OY 22 RVQNLKKADALCEVNESLIQODEKALYAAAGLOPKIAAAYAEQNFRTALSELASVK 81
 DB 582 RAFNLAEKVTHSV-LVDSLSFENNOEKALYQAISLE-----LTEDMHNDLKLFLALS 633
 OY 82 PQVDAFEDGVVMAEDAAYKONRLNLRLAEQNNAYADIALL 124
 DB 634 PIIDFEDNTWMTDDEKMKONRLAILNSIVAKARTVAAFNML 676
 RESULT 9
 ABP07415
 ID ABP07415 standard; Protein: 80 AA.
 XX
 AC ABP07415;
 XX
 DT 24-JUN-2002 (first entry)
 XX
 DE Human ORFX protein sequence SEQ ID NO:14812.
 XX
 KM Human; open reading frame; ORFX; gene therapy; cancer; cirrhosis;
 KM hyperproliferative disorder; psoriasis; benign tumour; haemorrhage;
 KM degenerative disorder; osteoarthritis; neurodegenerative disorder;
 KM cardiovascular disease; diabetes mellitus; systemic lupus erythematosus;
 KM hypertension; hypothyroidism; cholesterol ester storage disease;
 KM immune deficiency; immune disorder; infectious disease;
 KM autoimmune disorder; rheumatoid arthritis; autoimmune thyroiditis;
 KM myasthenia gravis.
 XX
 OS Homo sapiens.
 XX
 PN WO200192523-A2.
 XX
 PD 06-DEC-2001.
 XX
 PF 29-MAY-2001; 2001WO-US10836.
 XX
 PR 30-MAY-2000; 2000US-206132P.
 PR 29-AUG-2000; 2000US-228716P.
 XX
 PA (CURA-) CURAGEN CORP.
 PA Shinkets RA, Leach MD;
 PI Shinkets RA, Leach MD;
 XX
 DR WPI; 2002-106308/14.
 DR N-PSDB; ABN23167.
 XX
 XX Novel human polypeptides and polynucleotides useful for diagnosing,
 PT preventing and treating cardiovascular disease, neurodegenerative,
 PT hyperproliferative disorders and autoimmune disorders -
 XX
 PS Disclosure; SEQ ID 14812; 1037pp; English.
 XX
 CC The present invention describes substantially purified human proteins
 CC (referred to as open reading frame, ORFX, where X is 1-11491 (see Table 1
 CC in the specification). ABN15762 to ABN27252 encode the human ORFX

expressed from the genome sequence are useful for raising specific antibodies, identification of L. monocytogenes and related organisms, and for biosynthesis and biodegradation, especially biosynthesis of Vitamin B12. The genome sequence and proteins encoded by it are also useful for selecting compounds that regulate gene expression and cell replication and modulate L. monocytogenes-related diseases. In addition, the genome sequence and proteins encoded by it are useful in pharmaceutical and vaccines compositions for the treatment or prevention of infections by L. monocytogenes and related organisms.

Note: The sequence data for this patent did not form part of the printed specification, but was obtained in electronic format directly from WIPO at ftp.wipo.int/pub/published_pct_sequences.

SQ Sequence 688 AA;

Query Match 17.3%; Score 103.5; DB 23; Length 688;

Best Local Similarity 29.8%; Pred. No. 0.025; Mismatches 43; Indels 7; Gaps 3;

Matches 31; Conservative 23; Mismatches 43; Indels 7; Gaps 3;

QY 19 ANKRVONLKKADALGEVNESLQODEEKALYAAAGLOPKTAA-AVNEGFRRLSEL 77

DB 585 ALSRYVKIKKKEDGV-EVDPALFENEYEQALFDKLEKLEKFDYAGLTITE-----RLKAF 638

QY 78 ASVKPQVDAFPDGVVMAEDAANKONRLNLNRLAEONNAVADI 121

DB 639 ADLRTTIDAYFDNTLYMSDNDLKNRRLALLPELASFIKEFQAM 682

RESULT 12

AAV27142

ID AAV27142 standard; protein; 1503 AA.

XX AAV27142;

XX 15-SEP-1999 (first entry)

XX Human slit-2 mature protein (Seq ID No: 6 of JP11164690).

XX Vertebrate-derived protein; slit protein; diagnosis; cancer; nerve;

XX muscle; endocrine system.

XX Homo sapiens.

XX JP11164690-A.

XX 22-JUN-1999.

XX 05-DEC-1997; 97JP-0335435.

XX 05-DEC-1997; 97JP-0335435.

XX (ASAH) ASAH KASEI KOGYO KK.

XX WPI; 1999-411830/35.

XX N-PSDB; AAX89162.

XX New vertebrate slit protein - useful for diagnosis and treatment of

XX cancers in nerves, muscle and endocrine system

XX Claim 5; Page 43-47; 102pp; Japanese.

XX The invention relates to a vertebrate-derived protein containing an

XX amino acid sequence shown in AAV27137 and AAV27139. The

XX vertebrate-derived protein has at least 55 % homology to one of sequences

XX shown in AAV27141-17143, and has slit protein-like activity. The

XX 60% homology to nucleic acid sequences AAX89161-163. The

XX vertebrate-derived proteins can be produced recombinantly by transforming

XX host cells with expression vectors comprising the encoding nucleic acids.

XX The proteins of the invention are for diagnosing and treating cancer of

XX the nerves, muscle and/or endocrine system.

XX Sequence 1503 AA;

Query Match 14.7%; Score 88; DB 20; Length 1503;
Best Local Similarity 25.0%; Pred. No. 3; Mismatches 63; Indels 8; Gaps 4;
Matches 33; Conservative 28; Mismatches 63; Indels 8; Gaps 4;

QY 1 LOAVAVFQOLPEAALAAANKRVONLKKADALGEVNESLQODE-EKALYAAAGLOP 59

DB 522 LEATGIFPKLPQURKINFNNKITTDEGAFGASGVNEILLTSNRLEWQHKMKRGLE- 580

QY 60 KTAAVAEGNEFTALSE-----LASVKPQVDAFPDGVVMAEDAANKONRLNLNRLAEQ 114

DB 581 SKTTLMLRSNRITTCVGNDSFGLSSVR-LLSLYDNOITTVAPGADPTLHSLTLNLAMP 639

QY 115 MNAVDIALIGE 126

DB 640 FMCNCTLAWLGE 651

RESULT 13

AAW96701

ID AAW96701 standard; protein; 1503 AA.

XX AAW96701;

XX 15-APR-1999 (first entry)

XX Slit-like protein amino acid sequence.

XX Slit-like polypeptide; diagnosis; treatment; nervous disease;

XX thyroid disease; adrenal disease; muscular disease.

XX Homo sapiens.

XX JP11018777-A.

XX 26-JAN-1999.

XX 09-JUL-1997; 97JP-0183683.

XX 09-JUL-1997; 97JP-0183683.

XX (ASAH) ASAH KASEI KOGYO KK.

XX WPI; 1999-161084/14.

XX New slit-like polypeptide - useful for diagnosis and treatment of

XX nervous, thyroid, adrenal and muscular diseases

XX Claim 1; Page 17-21; 50pp; Japanese.

XX The present sequence represents the mature protein sequence of

XX a slit-like polypeptide. The polypeptide is useful for the

XX diagnosis and the treatment of all nervous diseases, thyroid

XX diseases, adrenal diseases and muscular diseases.

XX Sequence 1503 AA;

XX Query Match 14.7%; Score 88; DB 20; Length 1503;

XX Best Local Similarity 25.0%; Pred. No. 3;

XX Matches 33; Conservative 28; Mismatches 63; Indels 8; Gaps 4;

QY 1 LOAVAVFQOLPEAALAAANKRVONLKKADALGEVNESLQODE-EKALYAAAGLOP 59

DB 522 LEATGIFPKLPQURKINFNNKITTDEGAFGASGVNEILLTSNRLEWQHKMKRGLE- 580

QY 60 KTAAVAEGNEFTALSE-----LASVKPQVDAFPDGVVMAEDAANKONRLNLNRLAEQ 114

DB 581 SKTTLMLRSNRITTCVGNDSFGLSSVR-LLSLYDNOITTVAPGADPTLHSLTLNLAMP 639

QY 115 MNAVDIALIGE 126

DB 640 FMCNCTLAWLGE 651

RESULT 14
 AAY17499
 ID AAY17499 standard; Protein; 1525 AA.
 AC AAY17499;
 XX
 XX 04-AUG-1999 (first entry)
 DT
 XX Human Slt-1 protein.
 DE
 XX Human; Slt-1; Robo; modulation; identification; interaction.
 KW
 XX Homo sapiens.
 OS
 XX MO9925831-A2.
 PN
 XX 27-MAY-1999.
 PD
 XX 13-NOV-1998; 98MO-0524245.
 PF
 XX 07-APR-1998; 98US-0081057.
 PR 14-NOV-1997; 97US-0065544.
 XX
 XX (REGC) UNIV CALIFORNIA.
 PA
 XX Brose K, Goodman C, Kid T, Tessier-Lavigne M;
 PI
 XX WPI; 1999-347476/29.
 XX DR N-PSDB; AAX76161.
 XX
 XX Human Slt1 polypeptide and related nucleic acids
 PS Disclosure; Page 19-21; 34pp; English.
 XX
 XX The present sequence is the human Slt-1 protein. The present invention
 CC also describes a method for identifying agents which modulate the
 CC interaction of Robo and a Robo ligand comprising: combining a Robo
 CC polypeptide, a Slt1 polypeptide and a candidate agent under conditions
 CC where the Robo and Slt1 polypeptides normally (but for the presence of
 CC the agent) engage in a first interaction, where the Slt1 polypeptide
 CC specifically binds, activated or inhibits the activation of the Robo
 CC polypeptide and determining a second interaction of the Robo and Slt1
 CC polypeptides in the presence of the agent, where a difference between
 CC the first and second interactions indicates that the agent modulates the
 CC interaction of the Robo and Slt1 polypeptides; and a method to modulate
 CC the interaction of Robo and a Robo ligand. The method is useful for
 CC screening for Robo (roundabout) modulators and Robo:Slt1 complexes are
 CC useful for regulating various cell functions, especially of neuronal
 CC cells.
 XX
 XX Sequence 1525 AA;
 SQ
 Query Match. 14.7%; Score 88; DB 20; Length 1525;
 Best Local Similarity 25.0%; Pred. No. 3;
 Matches 33; Conservative 28; Mismatches 63; Indels 8; Gaps 4;
 OY 1 IQAVAVFKQLPEPAALAAANKRVONILKKADALGEVNESLIQODE-EKALYAAAGIOP 59
 DB 544 LEATGIFKRLPOLRKINFSNNKITDIEGAFEGASGVNEILLTSNRLNVOHKMKFGLE- 602
 OY 60 KITAAVAEGNFRITALSE-----LASVKPOVDAFEDGVWMAEDDAVKONRLNLRLAEQ 114
 DB 603 SLKTLMLRSNRITCVGNDSFGLSSVR-LLSLYDNOITTVAPGAFDTLHSLSTLNLAMP 661
 OY 115 MNAVADIALGE 126
 DB 662 FNCNCTYLAWLGE 673

RESULT 15
 AAY27145
 ID AAY27145 standard; protein; 1529 AA.

XX
 AC AAY27145;
 XX
 XX 15-SEP-1999 (first entry)
 DT
 XX Human silt-2 protein (Seq ID NO: 12 of JP11164690).
 DE
 XX Vertebrate-derived protein; silt protein; diagnosis; cancer; nerve;
 KW muscle; endocrine system.
 XX
 XX Homo sapiens.
 OS
 XX JP11164690-A.
 PN
 XX 22-JUN-1999.
 PD
 XX 05-DEC-1997; 97JP-0335435.
 PF
 XX 05-DEC-1997; 97JP-0335435.
 PR
 XX (ASAH) ASAH KASEI KOGYO KK.
 PA
 XX WPI; 1999-411830/35.
 DR N-PSDB; AAX89162.
 XX
 XX New vertebrate silt protein - useful for diagnosis and treatment of
 PT cancers in nerves, muscle and endocrine system
 PS Disclosure; Page 58-63; 102pp; Japanese.
 XX
 XX The invention relates to a vertebrate-derived protein containing an
 CC amino acid sequence shown in AAY27137 and AAY27139. The
 CC vertebrate-derived protein has at least 55 % homology to one of sequences
 CC shown in AAY27141-127143, and has silt protein-like activity. The
 CC vertebrate silt proteins encoding nucleic acid sequences have at least
 CC 60% homology to nucleic acid sequences AAX89161-163. The
 CC vertebrate-derived proteins can be produced recombinantly by transforming
 CC host cells with expression vectors comprising the encoding nucleic acids.
 CC The proteins of the invention are for diagnosing and treating cancer of
 CC the nerves, muscle and/or endocrine system.
 XX
 XX Sequence 1529 AA;
 SQ
 Query Match. 14.7%; Score 88; DB 20; Length 1529;
 Best Local Similarity 25.0%; Pred. No. 3;
 Matches 33; Conservative 28; Mismatches 63; Indels 8; Gaps 4;
 OY 1 IQAVAVFKQLPEPAALAAANKRVONILKKADALGEVNESLIQODE-EKALYAAAGIOP 59
 DB 548 LEATGIFKRLPOLRKINFSNNKITDIEGAFEGASGVNEILLTSNRLNVOHKMKFGLE- 606
 OY 60 KITAAVAEGNFRITALSE-----LASVKPOVDAFEDGVWMAEDDAVKONRLNLRLAEQ 114
 DB 607 SLKTLMLRSNRITCVGNDSFGLSSVR-LLSLYDNOITTVAPGAFDTLHSLSTLNLAMP 665
 OY 115 MNAVADIALGE 126
 DB 666 FNCNCTYLAWLGE 677

Search completed: December 2, 2002, 11:58:57
 Job time : 15.078 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 4.69005 Seconds
(without alignments)
790.458 Million cell updates/sec

Title: US-10-007-267-2

Perfect score: 598

Sequence: 1 LQAAVAFKOLPEAAALAAAN.....LLNRLAEQMAVADIALIGE 126

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/CTOS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	598	100.0	126	1 US-08-312-387B-2	Sequence 2, Appl1
2	598	100.0	126	1 US-08-683-426-2	Sequence 2, Appl1
3	598	100.0	126	1 US-08-683-458-2	Sequence 2, Appl1
4	598	100.0	126	2 US-08-878-360-2	Sequence 2, Appl1
5	598	100.0	126	3 US-08-478-140B-2	Sequence 2, Appl1
6	598	100.0	126	4 US-09-333-412-2	Sequence 2, Appl1
7	598	100.0	126	4 US-09-338-943-2	Sequence 2, Appl1
8	161	26.9	678	1 US-08-844-085-2	Sequence 2, Appl1
9	88	14.7	1525	3 US-09-191-647-2	Sequence 2, Appl1
10	88	14.7	1525	4 US-09-540-245A-2	Sequence 2, Appl1
11	88	14.7	1525	4 US-09-540-153-2	Sequence 2, Appl1
12	77	12.9	534	4 US-09-103-664A-2	Sequence 2, Appl1
13	75	12.5	1400	4 US-08-080-225-7	Sequence 7, Appl1
14	75	12.5	1400	5 US-08-465-713-7	Sequence 7, Appl1
15	75	12.5	1400	5 PCF-US93-05857-7	Sequence 7, Appl1
16	75	12.5	3969	4 US-08-061-376-5	Sequence 5, Appl1
17	74.5	12.5	3079	4 US-09-413-814-80	Sequence 80, Appl1
18	73.5	12.3	831	4 US-09-336-447A-1	Sequence 1, Appl1
19	72.5	12.1	373	4 US-09-071-035-116	Sequence 116, App
20	72.5	12.1	406	4 US-09-071-035-114	Sequence 114, App
21	72	12.0	92	4 US-09-344-529-2	Sequence 2, Appl1
22	71	11.9	610	4 US-09-336-447A-11	Sequence 11, Appl1
23	71	11.9	906	1 US-08-094-889-1	Sequence 8, Appl1
24	70	11.7	593	2 US-08-591-078-8	Sequence 10, Appl1
25	70	11.7	593	2 US-08-591-078-10	Sequence 9, Appl1
26	69.5	11.6	756	4 US-09-085-199B-9	Sequence 45, Appl1
27	69	11.5	660	4 US-09-268-347-45	

28	69	11.5	758	4 US-09-413-814-32	Sequence 32, Appl1
29	69	11.5	1111	1 US-08-317-450B-15	Sequence 15, Appl1
30	69	11.5	1111	4 US-08-800-593-15	Sequence 15, Appl1
31	69	11.5	1193	1 US-08-317-450B-13	Sequence 13, Appl1
32	69	11.5	1193	4 US-08-800-593-13	Sequence 13, Appl1
33	68.5	11.5	770	1 US-08-445-135-2	Sequence 2, Appl1
34	68.5	11.5	1151	3 US-08-840-006-6	Sequence 6, Appl1
35	68.5	11.5	1200	3 US-08-840-006-5	Sequence 5, Appl1
36	68	11.4	234	4 US-08-936-165A-517	Sequence 517, App
37	68	11.4	1261	3 US-09-080-855-2	Sequence 2, Appl1
38	68	11.4	1621	1 US-08-242-677-2	Sequence 2, Appl1
39	67.5	11.3	226	2 US-08-572-447C-15	Sequence 15, Appl1
40	67.5	11.3	226	4 US-09-267-747-15	Sequence 15, Appl1
41	67.5	11.3	289	1 US-08-072-070-4	Sequence 4, Appl1
42	67.5	11.3	289	1 US-08-469-434-4	Sequence 4, Appl1
43	67.5	11.3	289	1 US-08-214-222-4	Sequence 4, Appl1
44	67.5	11.3	289	2 US-08-467-852A-5	Sequence 5, Appl1
45	67.5	11.3	289	2 US-08-468-718-4	Sequence 4, Appl1

ALIGNMENTS

```
RESULT 1
US-08-312-387B-2
; Sequence 2, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauder & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-312-387B-2

Query Match 100.0%; Score 598; DB 1; Length 126;
Best Local Similarity 100.0%; Pred. No. 3.7e-57;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 LQAAVAFKOLPEAAALAAANKRYONILKKADALGVNSSLQODEKALYAAAGLOPK 60
|||||
DB 1 LQAAVAFKOLPEAAALAAANKRYONILKKADALGVNSSLQODEKALYAAAGLOPK 60
61 IAAVAFKOLPEAAALAAANKRYONILKKADALGVNSSLQODEKALYAAAGLOPK 120

Db 61 IAAVAEGNFRALSELASVYKPOVDAFFDGVVMAEDAAVKONRLNLRLAEOMNAVAD 120
OY 121 IALLGE 126
Db 121 IALLGE 126

RESULT 2

US-08-683-426-2

; Sequence 2, Application US/08683426

; Patent No. 5703567

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/683,426

; FILING DATE:

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/312,387

; FILING DATE: September 26, 1994

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-095B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201 487-5800

; TELEFAX: 201 343-1684

; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 126 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-683-426-2

Query Match 100.0%; Score 598; DB 1: Length 126;

Best Local Similarity 100.0%; Pred. No. 3.7e-57;

Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LOAVAVFKOLPEAALAAANKRVONLLKKADAALGEVNESLLQODEEKALYAAAGLOPK 60
Db 1 LOAVAVFKOLPEAALAAANKRVONLLKKADAALGEVNESLLQODEEKALYAAAGLOPK 60
OY 61 IAAVAEGNFRALSELASVYKPOVDAFFDGVVMAEDAAVKONRLNLRLAEOMNAVAD 120
Db 61 IAAVAEGNFRALSELASVYKPOVDAFFDGVVMAEDAAVKONRLNLRLAEOMNAVAD 120
OY 121 IALLGE 126
Db 121 IALLGE 126

RESULT 3

US-08-683-458-2

; Sequence 2, Application US/08683458

; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentin Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/683,458

; FILING DATE:

; CLASSIFICATION: 435

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/312,387

; FILING DATE: September 26, 1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-095A

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201 487-5800

; TELEFAX: 201 343-1684

; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 2:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 126 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

; US-08-683-458-2

Query Match 100.0%; Score 598; DB 1: Length 126;

Best Local Similarity 100.0%; Pred. No. 3.7e-57;

Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 LOAVAVFKOLPEAALAAANKRVONLLKKADAALGEVNESLLQODEEKALYAAAGLOPK 60
Db 1 LOAVAVFKOLPEAALAAANKRVONLLKKADAALGEVNESLLQODEEKALYAAAGLOPK 60
OY 61 IAAVAEGNFRALSELASVYKPOVDAFFDGVVMAEDAAVKONRLNLRLAEOMNAVAD 120
Db 61 IAAVAEGNFRALSELASVYKPOVDAFFDGVVMAEDAAVKONRLNLRLAEOMNAVAD 120
OY 121 IALLGE 126
Db 121 IALLGE 126

RESULT 4

US-08-878-360-2

; Sequence 2, Application US/08878360

; Patent No. 5945322

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/683,426
FILING DATE:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-0958
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-360-2

Query Match 100.0%; Score 598; DB 2; Length 126;
Best local Similarity 100.0%; Pred. No. 3.7e-57;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQAVAVFKQLPEAAALAAANKRVONLTKRADAALGEVNESLLOODEKALYAAAGLOPK 60
DB 1 LQAVAVFKQLPEAAALAAANKRVONLTKRADAALGEVNESLLOODEKALYAAAGLOPK 60
QY 61 IAAVAEGNFRRLSLASVKKPVDAFFDGVVYMAEDAANKRNLNRLAQMNAVAD 120
DB 61 IAAVAEGNFRRLSLASVKKPVDAFFDGVVYMAEDAANKRNLNRLAQMNAVAD 120
QY 121 IALIGE 126
DB 121 IALIGE 126

RESULT 5
US-08-478-140B-2
Sequence 2, Application US/08478140B
Patent No. 6127153
GENERAL INFORMATION:
APPLICANT: JOHNSON, KARL F.
APPLICANT: ROTH, STEPHEN
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Penite & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM PC compatible
COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,140B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENITE
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-140B-2

Query Match 100.0%; Score 598; DB 3; Length 126;
Best local Similarity 100.0%; Pred. No. 3.7e-57;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQAVAVFKQLPEAAALAAANKRVONLTKRADAALGEVNESLLOODEKALYAAAGLOPK 60
DB 1 LQAVAVFKQLPEAAALAAANKRVONLTKRADAALGEVNESLLOODEKALYAAAGLOPK 60
QY 61 IAAVAEGNFRRLSLASVKKPVDAFFDGVVYMAEDAANKRNLNRLAQMNAVAD 120
DB 61 IAAVAEGNFRRLSLASVKKPVDAFFDGVVYMAEDAANKRNLNRLAQMNAVAD 120
QY 121 IALIGE 126
DB 121 IALIGE 126

RESULT 6
US-09-333-412-2
Sequence 2, Application US/09333412
Patent No. 6342382
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: IBM floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800

```

      : TELEFAX: 1301 343-1684
      :       :
      : TELEX: 133521
      :
      : INFORMATION FOR SEQ ID NO: 2:
      :
      : SEQUENCE CHARACTERISTICS:
      :
      :     LENGTH: 126 amino acids
      :     TYPE: amino acid
      :     TOPOLOGY: linear
      :
      : MOLECULE TYPE: protein
      :
      : SEQUENCE DESCRIPTION: SEQ ID NO: 2:
      :

```

Query Match	100.0%	Score 598, DB 4	Length 126;
Best Local Similarity	100.0%	Pred. No. 3, 7e-57;	
Matches 126; Conservative	0;	Mismatches 0;	Gaps 0;

QY	1	LOAAVAFQLEBAAALAAANKRYONIKKDAALGEVNESLLOODEKALYYAAOGIOPK	60
QY	1	LOAAVAFQLEBAAALAAANKRYONILKKDAALGEVNESLLOODEKALYYAAOGIOPK <th>60</th>	60
Db	1	LOAAVAFQLEBAAALAAANKRYONILKKDAALGEVNESLLOODEKALYYAAOGIOPK <td>60</td>	60
QY	61	IAAAVABESNFTALSELASVAPQVDAPFDGVMVAEDAAVKONRLNLLNLAEDMNAVAD	120
Db	61	IAAAVABESNFTALSELASVAPQVDAPFDGVMVAEDAAVKONRLNLLNLAEDMNAVAD	120

QY	121	IALLGE	126
Db	121	IALLGE	126

RESULT 7
US-09-338-943-2

GENERAL INFORMATION:
APPLICANT: JOHNSON, KARL F.
APPLICANT: ROTH, STEPHEN
APPLICANT: BOCCALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:

```

?      COMPUTER READABLE FORM:
?      MEDIUM TYPE: Floppy disk
?      COMPUTER: IBM PC compatible
?      OPERATING SYSTEM: PC-DOS/MS-DOS
?      SOFTWARE: PatentIn Release #1.0, Version #1.30
?      CURRENT APPLICATION DATA:
?      APPLICATION NUMBER: US/09/338,943
?

```

REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-0177
TELECOMMUNICATION INFORMATION:

```

; INFORMATION FOR SEQ ID NO: 2:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 126 amino acids

```

4.

MOLECULE TYPE: protein
US-09-338-943-2

Query Match	100.0%;	Score 598;	DB 4;	Length 126;
Best Local Similarity	100.0%;	Pred. No. 3.7e-57;		
Matches 126;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY 1 LQAAVAFQJDEEAAALIAANKRYONILK KDDAALGEVNESITLQODEKALYAAAGCLOPK 60

Db 1 LQAAVAFQJDEEAAALIAANKRYONILK KDDAALGEVNESITLQODEKALYAAAGCLOPK 60

QY 61 IAAAVBESNERTALSELASVQYDAFFDGYMVAEDPAAYKQKRLNLINLTLAQMNAVAD 120

Db 61 IAAAVBESNERTALSELASVQYDAFFDGYMVAEDPAAYKQKRLNLINLTLAQMNAVAD 120

QY	121	IALLGE	126
Db	121	IALLGE	126

RESULT 8
US-08-844-085-2

GENERAL INFORMATION:
APPLICANT: Lawlor, Elizabeth
TITLE OF INVENTION: No. 5756330e1 Compounds
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: smitKline Beecham Corporation
STREET: 709 Swedeland Road
CITY: King of Prussia
STATE: PA

```

1  COMPUTER READABLE FORM:
2  MEDIUM TYPE: Diskette
3  COMPUTER: IBM Compatible
4  OPERATING SYSTEM: DOS
5  SOFTWARE: FastSoft for Windows Version 2.0.
6  CURRENT APPLICATION DATA:
7  APPLICATION NUMBER: US-/08/844.085
8  FILING DATE: 18-APR-1997

```

TELECOMMUNICATION INFORMATION:
TELEPHONE: 610-270-4478
TELEFAX: 610-270-5090

```

; INFORMATION FOR SEQ ID NO: 2
; SEQUENCE CHARACTERISTICS:
; LENGTH: 678 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
;
US-08-844-085-2

```

Query Match	26.98;	Score 161;	DB 1;	Length 678;
Best Local Similarity	35.58;	Pred. No. 3.5e-09;		
Matches	43;	Conservative	21;	Mismatches 37;
			Indels	20;
			Gaps	3

Dy 6 VERQLPEAALAAANK-----RQNLKRAADALGEVNSLSLQDDEEKALAAA 53
| | :|| :| | :|| :| :| :| :| :| :|
Db 553 VADMLEASALVEYSKEEDFKPSVESLSRAFNLAEKAE-GVATVDSALFENDOEKALAEA 611

QY 54 A QGLQPKIAA VAEGN FRTALSELASVKPQVDAF FDGVMMAEDA VKQNRNLNLRLAE 113

;; CURRENT FILING DATE: 1998-06-23
;; NUMBER OF SEQ ID NOS: 10
;; SOFTWARE: PatentIn version 3.0
;; SEQ ID NO 2
;; LENGTH: 534
;; TYPE: PRF
;; ORGANISM: Streptococcus equi
US-09-103-664A-2

Query Match 12.9%; Score 77; DB 4; Length 534;
Best Local Similarity 24.7%; Pred. No. 2.8;
Matches 37; Conservative 21; Mismatches 40; Indels 52; Gaps 6;

Qy 2 QAVAVFKQLPE--AAALAAANKRVONT-----LKKADA 32
Db 260 KAFVSKDLAKLAKTAEAEKIMENVSGLRLVESAKREMAQKLAETDQLTADKAKADA 319
Qy 33 ALGEVNESL--LQODEFALYAAAGLOPKTAAVAEGNFRTALSELVSPQVDAFTDG 90
Db 320 ELAANDPTIASLQTELEKRA-----KTELAVSERLIESGKREIAELQKRD----- 364
Qy 91 VVWVEDAAVKNRNLNLRLAEQMAVAD 120
Db 365 ----ASDKALVESQANVAE--LEKQKASD 388

RESULT 13

US-08-080-255-7
; Sequence 7, Application US/08080255
; Patent No. 5487970
; GENERAL INFORMATION:
; APPLICANT: Rowley, Janet D.
; APPLICANT: Diaz, Manuel O.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND
; TITLE OF INVENTION: TRANSLOCATIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/080.255
; FILING DATE: 19930617
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: AMINO ACID
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-080-255-7

Query Match 12.5%; Score 75; DB 1; Length 1400;
Best Local Similarity 32.6%; Pred. No. 17;
Matches 29; Conservative 13; Mismatches 25; Indels 22; Gaps 4;

Qy 5 AVFKQLEPAALAAANKRVONLLKKADALGEV--NESLQODEKALYAAAGLO----- 58
Db 227 SIGSMLAQADKLPMTDKRVASLLKRAKAQCKIEKSKSLKOTDQPK-----AQGESDSS 281
Qy 59 -----PKI-----AAVAEGNFRTALSE 76
Db 282 ETSVGRPRIKHVCRRAVAALGRKRAVFPD 310

RESULT 14
US-08-465-713-7
; Sequence 7, Application US/08465713
; Patent No. 6121419
; GENERAL INFORMATION:
; APPLICANT: Rowley, Janet D.
; APPLICANT: Diaz, Manuel O.
; TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR
; TITLE OF INVENTION: DETECTING GENE REARRANGEMENTS AND
; TITLE OF INVENTION: TRANSLOCATIONS
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Arnold, White & Durkee
; STREET: P. O. Box 4433
; CITY: Houston
; STATE: Texas
; COUNTRY: USA
; ZIP: 77210
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/465.713
; FILING DATE: 06-JUN-1995
; CLASSIFICATION: 530
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US/08/080.255
; FILING DATE: 17 JUNE 1993
; ATTORNEY/AGENT INFORMATION:
; NAME: Parker, David L.
; REGISTRATION NUMBER: 32,165
; REFERENCE/DOCKET NUMBER: ARCD:072/PAR
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (512) 320-7200
; TELEFAX: (512) 474-7577
; INFORMATION FOR SEQ ID NO: 7:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 1400 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: DNA (genomic)
US-08-465-713-7

Query Match 12.5%; Score 75; DB 3; Length 1400;
Best Local Similarity 32.6%; Pred. No. 17;
Matches 29; Conservative 13; Mismatches 25; Indels 22; Gaps 4;

Qy 5 AVFKQLEPAALAAANKRVONLLKKADALGEV--NESLQODEKALYAAAGLO----- 58
Db 227 SIGSMLAQADKLPMTDKRVASLLKRAKAQCKIEKSKSLKOTDQPK-----AQGESDSS 281
Qy 59 -----PKI-----AAVAEGNFRTALSE 76
Db 282 ETSVGRPRIKHVCRRAVAALGRKRAVFPD 310

RESULT 15
PCT-US93-05857-7
; Sequence 7, Application PC/TUS9305857
; GENERAL INFORMATION:
; APPLICANT: Board of Regents

APPLICANT: The University of Texas System
TITLE OF INVENTION: COMPOSITIONS AND METHODS FOR DETECTING
TITLE OF INVENTION: GENE REARRANGEMENTS AND TRANSLOCATIONS
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Arnold, White & Durkee
STREET: P. O. Box 4433
CITY: Houston
STATE: Texas
COUNTRY: USA
ZIP: 77210
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US93/05857
FILING DATE: 19930617
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/900,689
FILING DATE: 17/06/92
ATTORNEY/AGENT INFORMATION:
NAME: Parker, David L.
REGISTRATION NUMBER: 32,165
REFERENCE/DOCKET NUMBER: ARCD:072/PAR
TELECOMMUNICATION INFORMATION:
TELEPHONE: (512) 320-7200
TELEFAX: (512) 474-7577
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1400 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
PCT-US93-05857-7

Query Match 12.5%; Score 75; DB 5; Length 1400;
Best Local Similarity 32.6%; Pred. No. 17;
Matches 29; Conservative 13; Mismatches 25; Indels 22; Gaps 4;
QY 5 AVFKQLPEAAALAAANKRYONILTKKADALGEV--NESLIQDEEKALYAAAGLQ----- 58
DB 227 STGSMLAQADKLPMTDKRVASLKKAKAQLCKIEKSKSLKQTDQPK-----AQGESDSS 281
QY 59 -----PKT-----AAVAEGNFTALSE 76
DB 282 ETSVRCPRITKHVCRAAVALGRRAVFPD 310

Search completed: December 2, 2002, 12:03:12
Job time : 6.69005 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 12:00:15 ; Search time 2.52541 Seconds
(without alignments)
794.504 Million cell updates/sec

Title: US-10-007-267-2

Perfect score: 598
Sequence: 1 LOAAVAFKQLEPEAAALAAAN.....ILNRLAEQMAVADIALIGE 126

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCF_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_NEW_PUB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PCFS_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	598	100.0	126	US-10-007-267-2	Sequence 2, Appl1
2	163	27.3	989	US-09-815-242-4897	Sequence 4897, Ap
3	77.5	13.0	764	US-09-815-242-5143	Sequence 5143, Ap
4	75.5	12.6	271	US-09-847-539A-10	Sequence 10, Appl
5	75.5	12.6	1289	US-09-712-363-259	Sequence 259, Appl
6	75	12.5	905	US-09-905-963-58	Sequence 58, Appl
7	74	12.4	2186	US-09-815-242-12913	Sequence 12913, A
8	73.5	12.3	400	US-09-881-752A-206	Sequence 206, App
9	73.5	12.3	550	US-09-815-242-10499	Sequence 10499, A
10	73	12.2	2025	US-09-815-242-5703	Sequence 5703, Ap
11	73	12.2	3158	US-09-815-242-12611	Sequence 12611, A
12	72	12.0	393	US-09-815-242-11856	Sequence 11856, A
13	71.5	12.0	2437	US-09-815-242-5834	Sequence 5834, Ap
14	71.5	12.0	6281	US-09-815-242-12996	Sequence 12996, A
15	71	11.9	259	US-09-847-539A-8	Sequence 8, Appl1
16	70	11.7	2434	US-09-815-242-5835	Sequence 5835, Ap
17	70	11.7	2478	US-09-815-242-5816	Sequence 5816, Ap
18	69.5	11.6	2478	US-09-815-242-12967	Sequence 12967, A
19	69.5	11.6	878	US-10-108-605-325	Sequence 325, App

20	69.5	11.6	1043	US-09-946-805-4	Sequence 4, Appl1
21	69	11.5	635	US-09-815-242-11209	Sequence 11209, A
22	69	11.5	1111	US-09-756-071B-15	Sequence 15, Appl
23	69	11.5	1193	US-09-756-071B-13	Sequence 13, Appl
24	69	11.5	1215	US-09-815-242-5908	Sequence 5908, Ap
25	69	11.5	1269	US-09-815-242-13113	Sequence 13113, A
26	69	11.5	2368	US-09-815-242-5635	Sequence 5635, Ap
27	69	11.5	2368	US-09-815-242-12389	Sequence 12389, A
28	68.5	11.5	355	US-09-741-669-435	Sequence 435, App
29	68.5	11.5	605	US-09-841-132-574	Sequence 574, App
30	68	11.4	234	US-09-939-960-517	Sequence 517, App
31	68	11.4	861	US-09-815-242-14088	Sequence 14088, A
32	68	11.4	1523	US-10-052-586-290	Sequence 290, App
33	67.5	11.3	266	US-10-067-989-3	Sequence 3, Appl1
34	66.5	11.1	592	US-10-047-593-2	Sequence 2, Appl1
35	66.5	11.1	592	US-10-047-593-4	Sequence 4, Appl1
36	66.5	11.1	874	US-09-815-242-11127	Sequence 11127, A
37	66.5	11.1	891	US-09-815-242-10147	Sequence 10147, A
38	66.5	11.1	903	US-09-905-983-56	Sequence 56, Appl
39	66	11.0	892	US-09-815-242-13843	Sequence 13843, A
40	65.5	11.0	577	US-09-815-242-13752	Sequence 13752, A
41	65.5	11.0	651	US-09-841-132-405	Sequence 405, App
42	65.5	11.0	651	US-10-007-693-78	Sequence 78, Appl
43	65.5	11.0	874	US-09-815-242-11809	Sequence 11809, A
44	65	10.9	303	US-09-925-297-740	Sequence 740, App
45	65	10.9	891	US-09-272-809-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1
US-10-007-267-2
Sequence 2, Application US/10007267
Patent No. US20020127682A1
GENERAL INFORMATION:
APPLICANT: Gottschlich, Emil C.
TITLE OF INVENTION: OLIGOSYLTRANSFERASES FOR BIOSYNTHESIS OF GLYCOSYLCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 2:
SEQUENCE CHARACTERISTICS:
LENGTH: 126 amino acids
TYPE: amino acid

```

;
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 2
US-10-007-267-2

Query Match      100.0%; Score 598; DB 12; Length 126;
Best Local Similarity 100.0%; Pred. No. 8.2e-54;
Matches 126; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 LQAVAFKQLEPEAALAAANKRVONLTKKADPAALGEVNESILQODEEKALYAAAGLQPK 60
D 1 LQAVAFKQLEPEAALAAANKRVONLTKKADPAALGEVNESILQODEEKALYAAAGLQPK 60
QY 61 IAAVAEGNFRTALSELASVKPQVDAPFDGVVMAEDAAYKONRLNLRLAEQMAVAD 120
D 61 IAAVAEGNFRTALSELASVKPQVDAPFDGVVMAEDAAYKONRLNLRLAEQMAVAD 120
QY 121 IALGE 126
D 121 IALGE 126

RESULT 2
US-09-815-242-4897
; Sequence 4897, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 4897
; LENGTH: 989
; TYPE: PRP
; ORGANISM: Enterococcus faecalis
; FEATURE:
; NAME/KEY: VARIANT
; LOCATION: (1)...(989)
; OTHER INFORMATION: Xaa = Any Amino Acid
US-09-815-242-4897

Query Match      27.3%; Score 163; DB 10; Length 989;
Best Local Similarity 31.5%; Pred. No. 1.1e-08;
Matches 46; Conservative 19; Mismatches 55; Indels 26; Gaps 3;

QY 1 LQAVAFKQLEPEAALAAANKRVONLTKKADPAALGEVNESILQODEEKALYAAAGLQPK 39
D 848 IDAVVSAEQADLSKLPASANIILKSPEDDDFKPSMEALTRVINLAKGGLDTEEGCID 907
```

```

QY 40 -SLIQODEEKALYAAAGLQPKIAAVAEGNFRTALSELASVKPQVDAPFDGVVMAEDA 98
D 908 PSLEENNAEKELYOAVNDLSAPAFRTIAENYEA---LVNRLPIDAYFNETVMWDE 963
QY 99 AVKONRLNLRLAEQMAVADIAL 124
D 964 KVKONRLKQIMQIMKALMSIASIDL 989

RESULT 3
US-09-815-242-5143
; Sequence 5143, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 5143
; LENGTH: 764
; TYPE: PRP
; ORGANISM: Pseudomonas aeruginosa
US-09-815-242-5143

Query Match      13.0%; Score 77.5; DB 10; Length 764;
Best Local Similarity 30.2%; Pred. No. 3.4;
Matches 42; Conservative 17; Mismatches 53; Indels 27; Gaps 7;

QY 2 QAVAFKQLEPEAALAAANKRVONLTKKADPAALGEVNESILQODEEKALYAAAGLQPK 58
D 209 QAOQVTOQLGADENALTTPATK-VQQRKAYDASLSQYSTQV-----SGAVASATQANQ 262
QY 59 PKIAAVAEGN-----FRT--ALSELASVKPQVDAPFDGV-VMAEDAAYKQ 102
D 263 VKTAQAQQAQNSQISSQONLISAAPFNQJALAAQASGEAQTAFNQVQVIVEATFAFST 322
QY 103 NRNLNLRLAEQMAVADI 121
D 323 ARKQALAOFA--AMAAAEV 339

RESULT 4
US-09-847-539A-10
; Sequence 10, Application US/09847539A
; Patent No. US20020061306A1
; GENERAL INFORMATION:
; APPLICANT: Bjorck, Lars H
```

```

; APPLICANT: Rassmussen, Magnus
; TITLE OF INVENTION: STREPTOCOCCAL ALPHA ZM BINDING PROTEIN
; FILE REFERENCE: 100084.415US / N.75312B
; CURRENT APPLICATION NUMBER: US/09/847,539A
; CURRENT FILING DATE: 2001-05-01
; NUMBER OF SEQ ID NOS: 30
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 10
; LENGTH: 271
; TYPE: PRT
; ORGANISM: Streptococcus pyogenes
US-09-847-539A-10
```

```

Query Match          12.6%; Score 75.5; DB 10; Length 271;
Best Local Similarity 27.5%; Pred. No. 1.4;
Matches 33; Conservative 16; Mismatches 40; Indels 31; Gaps 5;
```

```

QY 11 PEAAALAAANKRVQNLKKA-----DALGEVNESLLQODEKALYAAAGCLQPK 60
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 28 PEKIALRNEERAIDELKKAIEDKEATTAIEAASSDALEALADQ-----ADALQSE 78
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 61 IAAVAEGFRRLASLAVKPOVDAFFDGVWVMAEDAAVKQNRLLNLRLAEQNNAVAD 120
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 79 EAAVQSDN--AASDALALADQADA-----LQSEAAVQSD---NAAGDALEALAD 126
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```

RESULT 5
US-09-712-363-259
; Sequence 259, Application US/09712363
; Patent No. US20020164588A1
; GENERAL INFORMATION:
; APPLICANT: Eisenberg, David
; APPLICANT: Rotstein, Sergio H.
; APPLICANT: Marcotte, Edward M.
; TITLE OF INVENTION: DETERMINING THE FUNCTIONS AND
; FILE REFERENCE: 07419-032001
; CURRENT APPLICATION NUMBER: US/09/712,363
; CURRENT FILING DATE: 2000-11-13
; PRIOR APPLICATION NUMBER: PCT/US00/02246
; PRIOR FILING DATE: 2000-01-28
; PRIOR APPLICATION NUMBER: 60/179,531
; PRIOR FILING DATE: 2000-02-01
; PRIOR APPLICATION NUMBER: 60/117,844
; PRIOR FILING DATE: 1999-01-29
; PRIOR APPLICATION NUMBER: 60/118,206,
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: 60/126,593
; PRIOR FILING DATE: 1999-03-26
; PRIOR APPLICATION NUMBER: 60/134,093
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/134,092
; PRIOR FILING DATE: 1999-05-14
; PRIOR APPLICATION NUMBER: 60/165,124
; PRIOR FILING DATE: 1999-11-12
; PRIOR APPLICATION NUMBER: 60/165,086
; PRIOR FILING DATE: 1999-11-12
; NUMBER OF SEQ ID NOS: 292
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 259
; LENGTH: 1289
; TYPE: PRT
; ORGANISM: Mycobacterium tuberculosis
US-09-712-363-259
```

```

Query Match          12.6%; Score 75.5; DB 9; Length 1289;
Best Local Similarity 25.0%; Pred. No. 11;
Matches 31; Conservative 21; Mismatches 65; Indels 7; Gaps 2;
```

```

QY 10 LPEAALAAANKRVQNLKKA-----AALGEVNESLLQODEKALYAAAGCLQPKIAA 63
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 245 IEBAAGVLRHRRKREKARLDTMAANLRLDTLTRELROLKPIGRQAEAAQRAAIAQA 304
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```

QY 64 AVAEGNFRRLASLAVKPOVDAFFDGVWVMAEDAAVKQNRLLNLRLAEQNNAVADIA 122
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 305 DLRLDALRLAADLVSRRAERREAVFQAEAAARRREHDEAAARLAVASBELAAHSAVAELS 364
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 123 LIGE 126
    |
DB 365 TRAE 368
```

```

RESULT 6
US-09-905-983-58
; Sequence 58, Application US/09905983
; Patent No. US20020045591A1
; GENERAL INFORMATION:
; APPLICANT: Geiger, Benjamin
; APPLICANT: Ben-Ze'ev, Avri
; APPLICANT: Sadot, Elnat
; TITLE OF INVENTION: METHODS AND THERAPEUTIC COMPOSITIONS FOR TREATING CANCER
; FILE REFERENCE: 01/22326
; CURRENT APPLICATION NUMBER: US/09/905,983
; CURRENT FILING DATE: 2001-09-28
; NUMBER OF SEQ ID NOS: 60
; SOFTWARE: PatentIn version 3.1
; SEQ ID NO 58
; LENGTH: 905
; TYPE: PRT
; ORGANISM: Mus musculus
US-09-905-983-58
```

```

Query Match          12.5%; Score 75; DB 10; Length 905;
Best Local Similarity 24.0%; Pred. No. 7.7;
Matches 30; Conservative 21; Mismatches 56; Indels 18; Gaps 4;
```

```

QY 3 AAVFKQLEPEAALAAANKRVQNLKKAADALGEVNESLLQODEKA-----LYAAAG 56
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 223 ASQAFRLHPDVATRANRDYVFKQVDEAIAISAAQATSPTEAKGHGIGELAAALNE 282
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
QY 57 LQKI---AAVAEGNFRRLASLAVKPOVDAFFDGVWVMAEDAAVKQNRLLNLRLAE 113
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
DB 283 FDNKILDPMTTSEARFRSLSE-----RLSEITISGAALMADSCTRRD---REEMVA 333
    | | | | | | | | | | | | | | | | | | | | | | | | | | | |
```

```

QY 114 QNNAV 118
    | | |
DB 334 ECVNAV 338
```

```

RESULT 7
US-09-815-242-12913
; Sequence 12913, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Kari L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
```

```
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 12913
; LENGTH: 2186
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12913

Query Match
Best Local Similarity 25.8%; Score 74; DB 10; Length 2186;
Matches 32; Conservative 20; Mismatches 52; Indels 20; Gaps 6;

QY 1 LQAVAVFR-QLEPAAALAAANKRVONILKKADALGEVNESLQODEEKALYA-AAQGL 57
Db 2000 LKAITALKQDEELKARTADVDVLRKRNVALSDI-EVITREKESLRLIDINIAQOT 2058

QY 58 QPKIAAANAENFRTALSELASVKRQVDAPF-DGVVMAEDAAVKONRLINLRLAEQMN 116
Db 2059 YAKFRATVP-----EQLKAKVILIDQYVADGNRMIDEDAT-----LNDIKQHTQ 2103

QY 117 AVAD 120
Db 2104 FIVD 2107

RESULT 8
US-09-881-752A-206
; Sequence 206, Application US/09881752A
; Patent No. US20020115078A1
; GENERAL INFORMATION:
; APPLICANT: Kleantous, Harold
; APPLICANT: Al-Garawi, Amal
; APPLICANT: Miller, Charles
; APPLICANT: Tomb, Jean-Francois
; APPLICANT: Oomen, Raymond P.
; TITLE OF INVENTION: Identification of Polynucleotides
; TITLE OF INVENTION: Encoding No. US20020115078A1el Helicobacter Polypeptides in the
; FILE REFERENCE: 06132/041002
; CURRENT APPLICATION NUMBER: US/09/881,752A
; PRIOR FILING DATE: 2001-06-15
; PRIOR APPLICATION NUMBER: US 08/833,457
; NUMBER OF SEQ ID NOS: 370
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 206
; LENGTH: 400
; TYPE: PRT
; ORGANISM: Helicobacter pylori
US-09-881-752A-206

Query Match
Best Local Similarity 12.3%; Score 73.5; DB 10; Length 400;
Matches 30; Conservative 16; Mismatches 44; Indels 17; Gaps 4;

QY 24 ONILKKADALGEVNESLQ-----ODEEKALYAAAGLQ---PKIAAVALGGR 71
Db 128 QNLASNDVVLQAVFENLHSTLSKMSQLSQEKEKLNTQALKNSIQKISSIIDBOKTR 187

QY 72 TALSIELASVKPOVDAPFDGVVMAEDAAVKONRLINLRLAEQMN 118
Db 188 EV--TLKSITKTEODKL--ILSMQKDYAIYNRLTLLEKERQNLNL 229

RESULT 9
US-09-815-242-10499
; Sequence 10499, Application US/09815242
; Patent No. US20020061569A1
```

```
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO 10499
; LENGTH: 550
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-815-242-10499

Query Match
Best Local Similarity 12.3%; Score 73.5; DB 10; Length 550;
Matches 31; Conservative 21; Mismatches 43; Indels 41; Gaps 3;

QY 8 KQLEPAAALAAANKRVONILKKADA-----ALGEVNESLQODEEKALYA 52
Db 27 KXDEKATEKTSVKVYLVNLMENSEIGMSDIFTPQDEASINAGSNVFEGLYQDEKDLIP 86

QY 53 AAQGLQPKTAA-----AAEGNFRALSELASVKRQVDAPF----- 88
Db 87 AAKKEPEISEDGKRYTILKLRDGRKWSNGDAVTADEFAMRKLANPKQANVFFLLBST 146

QY 89 --DGVVMAEDAAVKO 102
Db 147 IINGTAITKEEKAPPE 162

RESULT 10
US-09-815-242-5703
; Sequence 5703, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA.011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
```



```
RESULT 13
US-09-815-242-5834
; Sequence 5834, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 5834
; LENGTH: 2437
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-5834

Query Match      12.0%; Score 71.5; DB 10; Length 2437;
Best Local Similarity 26.0%; Pred. No. 64;
Matches 34; Conservative 22; Mismatches 62; Indels 13; Gaps 4;

QY 5 AVEKQPEAAALAAANKRKYONLLKKADALGEVNESLQODEEKALYAAOGLOPKIAA- 63
DB 575 AVKQDTEQAQALDQMLDAIQSIADKDATRASAVYNAEPNKKQSYDEAVQNAESIAGL 634
QY 64 ---AAVEGNFRTALSELASVKKPOVDAFEDGVVMAEDAAYVKNRNLNLRL--AEQM--- 115
DB 635 NNPTINKGNVSSATQAVISSKNAL---DGVRLAQDKQTAGNSLNHLDLQLPAPAOQALE 690
QY 116 NAVADIALIGE 126
DB 691 NOINNATTRGE 701

RESULT 14
US-09-815-242-12996
; Sequence 12996, Application US/09815242
; Patent No. US20020061569A1
; GENERAL INFORMATION:
; APPLICANT: Haselbeck, Robert
; APPLICANT: Ohlsen, Karl L.
; APPLICANT: Zyskind, Judith W.
; APPLICANT: Wall, Daniel
; APPLICANT: Trawick, John D.
; APPLICANT: Carr, Grant J.
; APPLICANT: Yamamoto, Robert T.
; APPLICANT: Xu, H. Howard
; TITLE OF INVENTION: Identification of Essential Genes in
; FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; PRIOR FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 12996
; LENGTH: 6281
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996

Query Match      12.0%; Score 71.5; DB 10; Length 6281;
Best Local Similarity 26.0%; Pred. No. 2.3e+02;
Matches 34; Conservative 22; Mismatches 62; Indels 13; Gaps 4;
```

```
FILE REFERENCE: ELITRA 011A
; CURRENT APPLICATION NUMBER: US/09/815,242
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: 60/191,078
; PRIOR FILING DATE: 2000-03-21
; PRIOR APPLICATION NUMBER: 60/206,848
; PRIOR FILING DATE: 2000-05-23
; PRIOR APPLICATION NUMBER: 60/207,727
; PRIOR FILING DATE: 2000-05-26
; PRIOR APPLICATION NUMBER: 60/242,578
; PRIOR FILING DATE: 2000-10-23
; PRIOR APPLICATION NUMBER: 60/253,625
; PRIOR FILING DATE: 2000-11-27
; PRIOR APPLICATION NUMBER: 60/257,931
; PRIOR FILING DATE: 2000-12-22
; PRIOR APPLICATION NUMBER: 60/269,308
; PRIOR FILING DATE: 2001-02-16
; NUMBER OF SEQ ID NOS: 14110
; SOFTWARE: FastSeq for Windows Version 4.0
; SEQ ID NO: 12996
; LENGTH: 6281
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-815-242-12996

Query Match      11.9%; Score 71; DB 10; Length 259;
Best Local Similarity 27.5%; Pred. No. 3.7;
Matches 33; Conservative 17; Mismatches 42; Indels 28; Gaps 5;

QY 11 PEAALAAANKRKYONLLKRA-----DAALGEVNESLQODEEKALYAAOGLOPK 60
DB 45 PEKALRNERRAIDEIKKQIETKEATTAIEAASSALERLADQ-----ADALQSE 95
QY 61 IAAVAEGNFRTALSELASVKKPOVDAF--FDGVVMAEDAAYVKNRNLNLRLAEQMNAY 118
DB 96 EAAVOSDN--AASDALREALDQTDALQSEEAAYVADNNA-----SDTLEALADQTDAL 148
```

Search completed: December 2, 2002, 12:15:58
Job time : 5.52541 secs

THIS PAGE BLANK (USPTO)

A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-687 <PAR>
A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAR83815.1; PID:g737926
A:Experimental source: serogroup A, strain Z2491
C:Genetics:
A:Gene: g1ys; MMA0523
C:Superfamily: glycine-tRNA ligase beta chain
C:Keywords: ligase

Query Match 95.7%; Score 572; DB 2; Length 687;
Best Local Similarity 96.0%; Pred. No. 7.3e-40;
Matches 121; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

QY 1 LQAVNFKQLPEAALAAANKRVQNLKKRADAALGEVNESLLOODEKALVYAAAGLOPK 60
DB 564 LQAVNFKQLPEAALAAANKRVQNLKKRADAALGEVNESLLOODEKALVYAAAGLOPK 621
QY 61 IAAVAEGNFTALSELASVKPOVDAPFDGVVMAEDAAVKONRLNLRLAEQNNAVD 120
DB 622 IAAVAEGNFTALSELASVKPOVDAPFDGVVMAEDAAVKONRLNLRLAEQNNAVD 681
QY 121 IALLG 126
DB 682 IALLG 687

RESULT 3

S70812
glycine-tRNA ligase (EC 6.1.1.14) - Neisseria meningitidis (fragment)

N:Alternate names: glycyl tRNA synthetase
C:Species: Neisseria meningitidis
C:Date: 12-Feb-1998 #sequence_revision 20-Feb-1998 #text_change 03-Jun-2002
C:Accession: S70812
R:Jennings, M.P.; Hoog, D.W.; Peak, I.R.A.; Virji, M.; Moxon, E.R.
Mol. Microbiol. 18, 729-740, 1995
A:Title: Molecular analysis of a locus for the biosynthesis and phase-variable expression
A:Reference number: S70812; MUID:96414473; PMID:8817494
A:Accession: S70812
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-87 <GEN>
A:Cross-references: EMBL:U25839; NID:g9733183; PIDN:AAC44083.1; PID:g9733184
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C:Genetics:
A:Gene: g1ys
C:Superfamily: glycine-tRNA ligase beta chain
C:Keywords: ligase

Query Match 68.2%; Score 408; DB 2; Length 87;
Best Local Similarity 97.7%; Pred. No. 2.8e-27;
Matches 85; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

QY 40 SLLOODEKALVYAAAGLOPKIAAAGEGFRFALSELASVKPOVDAPFDGVVMAEDAA 99
DB 1 SLLOODEKALVYAAAGLOPKIAAAGEGFRFALSELASVKPOVDAPFDGVVMAEDAA 60
QY 100 VKONRLNLRLAEQNNAVADIALGE 126
DB 61 VKONRLNLRLAEQNNAVADIALGE 87

RESULT 4

F82373
glycyl-tRNA synthetase, beta chain VC0020 [imported] - Vibrio cholerae (strain N16961 se

C:Species: Vibrio cholerae
C:Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Feb-2001
C:Accession: F82373

R:Heidelberg, J.F.; Eisen, J.A.; Nelson, W.C.; Clayton, R.A.; Gwin, M.L.; Dodson, R.J.;
Chardon, D.; Ermolaeva, M.D.; Vamathevan, J.; Baas, S.; Qiu, H.; Dragoi, I.; Sellers, F.
1, R.R.; Mekalanos, J.J.; Venter, J.C.; Fraser, C.M.
Nature 406, 477-483, 2000

A:Title: DNA Sequence of both chromosomes of the cholera pathogen Vibrio cholerae.

A:Reference number: A82035; MUID:20406833; PMID:109522301
A:Accession: F82373
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-688 <HE>
A:Cross-references: GB:AE004094; GB:AE003852; NID:96654407; PIDN:AAF93198.1; GSPDB:GN
A:Experimental source: serogroup O1; strain N16961; biotype El Tor
C:Genetics:
A:Gene: VC0020
A:Map position: 1
C:Superfamily: glycine-tRNA ligase beta chain

Query Match 49.2%; Score 294.5; DB 2; Length 688;
Best Local Similarity 52.8%; Pred. No. 7.4e-17;
Matches 66; Conservative 23; Mismatches 35; Indels 1; Gaps 1;

QY 1 LQAVNFKQLPEAALAAANKRVQNLKKRADAALG-EVNESLLOODEKALVYAAAGLOPK 59
DB 564 VKAVSHFRLEAEALAAANKRVQNLKKRADAALGDELGEEDLALQDADKALAEAVEIMAE 621
QY 60 KIAAAGEGFRFALSELASVKPOVDAPFDGVVMAEDAAVKONRLNLRLAEQNNAVA 119
DB 622 ALEPAFATGNYOELSKLAGRAPVDAPFDGVVMAEDAEALKKRNLTLNKLRLDLQIA 681
QY 120 IALL 124
DB 682 IALL 686

RESULT 5

B64103
glycine-tRNA ligase (EC 6.1.1.14) beta chain - Haemophilus influenzae (strain Rd KW20

N:Alternate names: glycyl-tRNA synthetase beta chain
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 03-Jun-2002
C:Accession: B64103
R:Flieschman, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Goeyne, J.D.; Scott, J.; Shiley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: B64103
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-688 <TIGR>
A:Cross-references: GB:U32774; GB:U42023; NID:g1573942; PIDN:AAC22584.1; PID:g1573946
C:Superfamily: glycine-tRNA ligase beta chain
C:Keywords: aminoacyl-tRNA synthetase; ligase; protein biosynthesis

Query Match 45.7%; Score 273; DB 2; Length 688;
Best Local Similarity 47.6%; Pred. No. 4.5e-15;
Matches 59; Conservative 26; Mismatches 39; Indels 0; Gaps 0;

QY 1 LQAVNFKQLPEAALAAANKRVQNLKKRADAALGEVNESLLOODEKALVYAAAGLOPK 60
DB 564 VRAVSHFRFLDSAEALAAANKRVSNITLAKAALGIGINILACVEPEKALAEVIALKRE 623
QY 61 IAAVAEGNFTALSELASVKPOVDAPFDGVVMAEDAAVKONRLNLRLAEQNNAVD 120
DB 624 VQPLINQGDYTYLDKLANLRAPVDSFFDNVWMAEDPALRQRLAILMTIGSLFLQVAD 683
QY 121 IALL 124
DB 684 ISVL 687

RESULT 6

D83645

glycyl-tRNA synthetase beta chain PA0008 [imported] - Pseudomonas aeruginosa (strain

C:Species: Pseudomonas aeruginosa
C:Date: 15-Sep-2000 #sequence_revision 15-Sep-2000 #text_change 31-Dec-2000

A:	Reference number:	A64720;	MUID:97426617;	PMID:9278503
A:	Accession:	A65155		
A:	Status:	Preliminary:	nucleic acid sequence not shown; translation not shown	
A:	Molecule type:	DNA		
A:	Residues:	1-689 <BLAST>		
A:	Cross-references:	GB:AE000433; GB:U00096; NID:g1789977; PIDN:AAC76583.1; PID:g1789982		
A:	Experimental source:	strain K-12, substrain MG1655		
C:	Genetics:			
A:	Gene:	glys(b)		
A:	Map position:	80 min		
C:	Complex:	tetramer of two alpha and two beta chains		
C:	Superfamily:	glycine-CRNA ligase beta chain		
C:	Keywords:	aminoacyl-trna synthetase; ligase; protein biosynthesis		
Qy	Query Match	41.6%; Score 248.5; DB 1; Length 689;		
	Best Local Similarity	47.2%; Pred. No. 4.9e-13;		
	Matches 59; Conservative 21; Mismatches 44; Indels 1; Gaps 1;			
Qy	1	LOAVVAFKOLEEAAALAAANKRVQLTKKADAAAGC-VNESLLQDEKAKATYAAAQGLQP 59		
		::::: : :: : : : : : :		
Db	564	MKAVSHEFTLDAANAALAAANKRVSITLAKSDEVLSDRNAVASLTKEPEETKLAMQVVVLRL 623		
Qy	60	KIAAIVAEGNFRTAISLASVKPOVDAFFDGVYMAEDAAKQNPLNLINRLAEGMAYVA 119		
		: : : : : : : : : : : : : : :		
Db	624	KLEPFTGEGRYODALVELALEIREPDVAFDDKVMMVVDDEKRINRLTLMLEKLRDELFLRYVA 683		
Qy	120	DIALL 124		
		: : :		
Db	684	DISTLL 688		

C:Species: Salmonella enterica subsp. enterica serovar Typhimurium
A>Note: this species has also been called Salmonella typhi
C>Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #extL_change 03-Jun-2002
R:Accession: A10980
P:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church
ch, T.; Connerton, P.; Croplin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farr
S.; Moule, S.; O'Garra, P.
N:Nature 413, 848-852, 2001
A:Authors: Parry, C., Quail, M., Rutherford, K., Simmonds, M., Skelton, J., Stevens,
A.T.H.: Complete genome sequence of a multiple drug resistant Salmonella enterica se
A:Reference number: AB0502; PMID:11677608
A:Accession: AI0960
A>Status: preliminary
M:Molecule type: DNA
A:Residues: 1-689 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD07973.1; PID:g16504960; GSDB:GNO0176
C:Genetics:
A:Gene: STY4144
C:Superfamily: glycine-tRNA ligase beta chain
C:Keywords: Ligase

	Query Match	38.4%	Score 229..5;	DB 2;	Length 689;
	Best Local Similarity	44.8%	Pred. No. 1.8e-11;		
	Matches 56;	Conservative % 26;	Mismatches 42;	Indels 1;	Gaps 1;

OY 1 LGAAVAFKQPEBAALAANKRVONLKADALAGE-VNESLLOODEKKALYYAAAGLP 59
::: | : | ::|||::|:: | : | : | : | :
564 MKAVSHFTLEBASLAANKRRSVITLAKTEPLNDIVHASYVLKEAEIETLARHLVVLRD 623

RESULT 10
AH0494
glycine-tRNA ligase (EC 6.1.1.14) [imported] - Yersinia pestis (strain CO92)
C:Species: Yersinia pestis
C:Date: 02-Nov-2001 #sequence_revision 02-Nov-2001 #text_change 03-Jun-2002
C:Accession: AH0494
R:Parkhill, J., Wren, B.M.; Thomson, N.R.; Titball, R.W.; Holden, M.T.G.; Prentice, M.B.
deno-Tarrago, A.M.; Chillingworth, T.; Cronin, A.; Davies, R.M.; Davis, P.; Dougan, G.;
Lea, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens, K.; Whitehead, S.; Barrrell,
Nature 413, 523-527, 2001
A:Title: Genome sequence of Yersinia pestis, the causative agent of plague.
A:Reference number: AB0001; MUID:21470413; PMID:11586360
A:Accession: AH0494
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-689 <KUR>
A:Cross-references: GB:AL590842; PIDN:CAC93524.1; PID:g15981965; GSPDB:GN00175
C:Genetics:
A:Gene: g1yS
C:Superfamily: glycine-tRNA ligase beta chain
C:Keywords: ligase

```

Db      624  KQPFAGRGROYEALIELDALRLAPYDEFFENNVMNAEKRIDIRINRKLFLSKLRELFQVA 683
Oy      120  DIALL 124
        ||:||
Db      684  DISTL 688

RESULT 12
A82617
glycyl-tRNA synthetase beta chain XF1959 [imported] - Xylella fastidiosa (strain 9a5c
C.Species: Xylella fastidiosa
C.Date: 18-Aug-2000 #sequence_revision 20-Aug-2000 #text_change 02-Sep-2000
C.Accession: A82617
Randomus, The Xylella fastidiosa Consortium of the Organization for Nucleotide Seq
Nature 406, 151-157, 2000
A.Title: The genome sequence of the plant pathogen Xylella fastidiosa.
A.Reference number: A82515; MUID:20365717; PMID:10910347
A.Note: for a complete list of authors see reference number A59328 below
A.Accession: A82617
A.Status: preliminary
A.Molecule type: DNA

```

	Query Match Similarity	40.4%	Score 241.5:	DB 2:	Length 689:
	Best Local Similarity	46.4%	Pred No.18e-12:		
	Matches	58:	Conservative	23:	Mismatches 43: Indels 1: Gaps 1:
QY	1 LGAVVFVKLPPEAAALAAANKRVYNLLTKADALGE-VNESLQQDEKALYAAAGLIQP	59			
	:::: : : : : : :				
DB	564 VAAVYVFRLLDAAALAAANKRVSNTLAKSTDLNDHVAASITKEPAELKLATHLVLELD	623			
QY	60 KLAAPAAVEGSENFRTASELASVKKQVDADFQGVVMAEDDAVKONRLNLTLRLEQNNAVYA	119			
	:::: : : :				
DB	624 QLEPFAAQAQYKEALVELTAALRETVDDEFESVYVMAEDDAVRVNRRLTSLKURELFLOYA	683			
QY	120 DIALL 124				
	:::: :				
DB	684 DISTL 688				

A: Cross-references: GB:AE004015, GB:AE003849, NID:9107057, PIDN:AAF84761.1, GSPDB:GN
A: Experimental source: strain 9a5c
R: Simpson, A.J.G.: Reinach, F.C.; Arruda, P.; Abreu, F.A.; Acencio, M.; Alvarenga, R.;
Bittomes, M.R.S.; Bueno, M.R.P.; Camargo, A.A.; Camargo, L.E.A.; Carraro, D.M.; Carraro
da Neto, E.R.; Docena, C.; El-Dorriy, H.; Facincani, A.P.; Ferreira, A.J.S.
submitted to GenBank, June 2000
A: Authors: Ferreira, V.C.A.; Ferro, J.A.; Fraga, J.S.; Franca, S.C.; Franco, M.C.; Fr
J.D.; Junqueira, M.L.; Kemper, E.L.; Kitajima, J.P.; Klieger, J.E.; Kuxame, E.E.; La
Chado, M.A.; Madeira, A.M.B.N.; Madeira, H.M.F.; Marino, C.L.; Marques, M.V.; Martins
A: Authors: Martins, E.M.F.; Matsukuma, A.Y.; Menck, C.F.M.; Miranda, E.C.; Miyaki, C.;
F.G.; Nunes, L.R.; Oliveira, M.A.; de Oliveira, M.C.; de Oliveira, R.C.; Palmieri,
Rodrigues, V.; Rosa, A.J. de M.; de Rosa Jr., V.E.; de Sa, R.G.; Santelli, R.V.; Sawa
A: Authors: da Silva, A.C.R.; da Silva, F.R.; da Silva, A.M.; Silva Jr., W.A.; da Silv
M.; Tsuchiko, M.H.; Vallada, H.; Van Sluys, M.A.; Verjovski-Almeida, S.; Vettore, A.L.
A: Reference number: AS9328

RESULT 11
A10980
glycine-tRNA ligase (EC 6.1.1.14) - *Salmonella enterica* subsp. *enterica* serovar Typhi (S)

A:Gene: XF1959
C:Superfamily: glycine-tRNA ligase beta chain

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 3.06657 Seconds
(without alignments)
1704.189 Million cell updates/sec

Title: US-10-007-267-2

Perfect score: 598
Sequence: 1 LQAAVAFKQLEPEAAALAAAN.....LTNRLAEQMAVADIALIGE 126

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	598	100.0	126	SYGB_NEIGO	O50945 neisseria g
2	294.5	49.2	688	SYGB_VIBCH	O9KW8 vibrio chol
3	278	45.5	285	SYGB_HAEAE	O30836 haemophilus
4	273	45.7	688	SYGB_HAEIN	P43822 haemophilus
5	248.5	41.6	688	SYGB_ECOLT	P00961 escherichia
6	238	39.8	284	SYGB_COXBU	P45651 coxiella bu
7	231.5	36.7	689	SYGB_PASMO	P57905 pasteurella
8	210.5	35.2	722	SYGB_XYLFA	O9PC26 xyloella fas
9	151	25.3	672	SYGB_THEMA	O9WY60 thermotoga
10	138.5	23.2	665	SYGB_RICPR	O9ZCBI rickettsia
11	138	23.1	679	SYGB_BAGSU	P54381 bacillus su
12	137	22.9	664	SYGB_AQUAE	O67898 aquifex aco
13	137	22.9	693	SYGB_BARCH	O9KD48 bacillus ha
14	131.5	22.0	722	SYGB_SYNY3	O55690 synechocyst
15	130.5	21.8	690	SYGB_BUCAL	P57235 buchnera ap
16	86.5	14.5	810	HPR4_HALSA	O48317 halobacteri
17	84.5	14.1	189	PAAD_NEIMA	O9W78 neisseria m
18	83.5	14.0	810	HPR4_HALNI	O9HP84 halobacteri
19	82.5	13.8	579	G160_HUMAN	O08378 homo sapien
20	82.5	13.8	579	G160_HUMAN	O08378 homo sapien
21	82	13.7	500	MSP_DIRIM	P13392 dirofilaria
22	82	13.7	505	FLJB_SALAE	P52616 salmonella
23	81.5	13.6	879	MYSB_ONCYO	O02171 onchocerca
24	81.5	13.6	880	MYSB_ONCYO	O02171 onchocerca
25	80.5	13.5	189	PAAD_NEIMA	O01202 brugia mala
26	79.5	13.3	239	RNC_HELPY	O9YXP4 neisseria m
27	79.5	13.3	647	NAFL_MOUSE	P56118 heliobacte
28	79.5	13.3	882	MYSB_CAEEL	O9WU8 mus musculu
29	78.5	13.1	239	RNC_HELPY	O9ZKH2 mycobacteri
30	77.5	13.0	385	YL90_MYCTU	O10383 mycobacteri
31	77.5	13.0	636	NAFL_HUMAN	O15025 homo sapien
32	77	12.9	488	FLIC_SALCH	P06176 salmonella
33	77	12.9	493	FLIC_SALPA	P06178 salmonella

34	77	12.9	513	1	GRDC_EUBAC	P54935 eubacterium
35	77	12.9	701	1	SYGB_HELPY	P56454 heliobacte
36	76.5	12.8	572	1	Y062_DEIRA	O9TRM6 deinococcus
37	76.5	12.8	1115	1	TBC2_CHLRE	O8VXP3 chlamydomon
38	76.5	12.8	2477	1	SPCN_CHICK	P07751 gallus gall
39	76	12.7	701	1	SYGB_HELPY	O9ZKM9 heliobacte
40	75.5	12.6	1205	1	SMC_MYCTU	O10970 mycobacteri
41	75	12.5	196	1	EFYS_THETH	P43895 thermus the
42	75	12.5	953	1	CTN2_HUMAN	P26232 homo sapien
43	75	12.5	953	1	CTN2_HUMAN	O61301 mus musculu
44	75	12.5	1325	1	G160_MOUSE	P55937 mus musculu
45	75	12.5	1426	1	CUT2_MOUSE	P70298 mus musculu

ALIGNMENTS

RESULT 1	ID	SYGB_NEIGO	STANDARD:	PRT:	126 AA.
AC	O50945:				
DT	15-JUL-1998 (Rel. 36, Created)				
DT	15-JUL-1998 (Rel. 36, Last sequence update)				
DT	16-OCT-2001 (Rel. 40, Last annotation update)				
DE	Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase				
DE	beta chain) (GlyRS) (Fragment).				
GN	GLYS				
OS	Neisseria gonorrhoeae.				
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.				
OX	NCBI_TaxId=485;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=P62;				
RX	MEDLINE=95053752; PubMed=7964493;				
RA	Gotschlich E.C.;				
RT	"Genetic locus for the biosynthesis of the variable portion of				
RT	Neisseria gonorrhoeae lipooligosaccharide.";				
RL	J. Exp. Med. 180:2181-2190(1994).				
CC	- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) -> AMP + diphosphate				
CC	+ glycyl-tRNA(Gly).				
CC	- SUBUNIT: Tetramer of two alpha and two beta chains				
CC	(BY SIMILARITY).				
CC	- SUBCELLULAR LOCATION: Cytoplasmic.				
CC	- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.				
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration				
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -				
CC	the European Bioinformatics Institute. There are no restrictions on its				
CC	use by non-profit institutions as long as its content is in no way				
CC	modified and this statement is not removed. Usage by and for commercial				
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/				
CC	or send an email to license@isb-sib.ch).				
CC	-----				
DR	EMBL; U14554; AAA68008.1; -				
DR	PROSITE; P550861; AA-TRNA_LIGASE_II_GUYAB.1.				
KW	Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.				
FT	NON_TER				
FT	SEQUENCE 126 AA; 13343 MW; C3913AD1A6591CF2 CRC64;				
QY	Query Match	100.0%; Score 598; DB 1; Length 126;			
	Best Local Similarity	100.0%; Pred. No. 2.9e+43;			
	Matches 126; Conservative	0; Mismatches 0; Indels 0; Gaps 0;			
QY	1 LQAAVAFKQLEPEAAALAAANKRYONILTKRADAALGVENSLAQODEKALYAAAGLOPK 60				
DB	1 LQAAVAFKQLEPEAAALAAANKRYONILTKRADAALGVENSLAQODEKALYAAAGLOPK 60				
QY	61 LQAAVAFKQLEPEAAALAAANKRYONILTKRADAALGVENSLAQODEKALYAAAGLOPK 120				
DB	61 LQAAVAFKQLEPEAAALAAANKRYONILTKRADAALGVENSLAQODEKALYAAAGLOPK 120				
QY	121 IALLGE 126				

Db 121 IALLG 126

RESULT 2

SYGB_VIBCH STANDARD: PRT: 688 AA.

ID SYGB_VIBCH

AC 09KVM8;

DT 16-OCT-2001 (Rel. 40, Created)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase

DE beta chain) (GlyRS).

DE GLYS OR VC0020.

OS Vibrio cholerae.

OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.

OX NCBI_TaxID=666;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=E1 Tor M16961 / Serotype O1;

RX MEDLINE=20406833; PubMed=10952301;

RA Heidelberg J.F., Eisen J.A., Nelson W.C., Clayton R.A., Gwin M.L.,

RA Dodson R.J., Haft D.H., Hickey E.K., Peterson J.D., Umayam L.A.,

RA Gill S.R., Nelson K.E., Read T.D., Tettelin H., Richardson D.,

RA Ermolaeva M.D., Vamathevan J., Bass S., Qin H., Dragoi I., Sellers P.,

RA McDonald L., Uitterback T., Fleischmann R.D., Nierman W.C., White O.,

RA Salzberg S.L., Smith H.O., Colwell R.R., Mekalanos J.J., Venter J.C.,

RA Fraser C.M.;

RT "DNA sequence of both chromosomes of the cholera pathogen Vibrio

RT cholerae";

RL Nature 406:477-483(2000).

CC -1- CATALYTIC ACTIVITY: ATP + glycine + tRNA(gly) = AMP + diphosphate

CC + glycyl-tRNA(gly).

CC -1- SUBUNIT: Tetramer of two alpha and two beta chains

CC (by similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-tRNA SYNTHETASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL: AE004094; AAF93198.1; -

CC TIGR: VC0020; -

CC InterPro: IPR002311; trna_synth_2f.

CC Pfam: PF02092; trna_synth_2f.1.

CC PRINTS: PR01045; TRNASYNTHGB.

CC TIGRFAMs: TIGR00211; glyS; 1.

CC PROSITE: PS50861; AA_TRNA_LIGASE_IL_GLYAB; 1.

CC AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

RESULT 3

SYGB_HAEAE STANDARD: PRT: 285 AA.

ID SYGB_HAEAE

AC 030836;

DT 15-JUL-1998 (Rel. 36, Created)

DT 15-JUL-1998 (Rel. 36, Last sequence update)

DT 16-OCT-2001 (Rel. 40, Last annotation update)

DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase

DE beta chain) (GlyRS) (Fragment).

DE GLYS.

OS Haemophilus aegyptius.

OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;

OC Haemophilus.

OX NCBI_TaxID=725;

RN [1]

RP SEQUENCE FROM N.A.

RC STRAIN=F3031;

RL Tondella M.L.C., Utt E.A., Mayer L.W., Quinn F.D.;

CC Submitted (AUG-1997) to the EMBL/GenBank/DBJ databases.

CC -1- CATALYTIC ACTIVITY: ATP + glycine + tRNA(gly) = AMP + diphosphate

CC + glycyl-tRNA(gly).

CC -1- SUBUNIT: Tetramer of two alpha and two beta chains

CC (by similarity).

CC -1- SUBCELLULAR LOCATION: Cytoplasmic.

CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-tRNA SYNTHETASE FAMILY.

CC -----

CC This SWISS-PROT entry is copyright. It is produced through a collaboration

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

CC the European Bioinformatics Institute. There are no restrictions on its

CC use by non-profit institutions as long as its content is in no way

CC modified and this statement is not removed. Usage by and for commercial

CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>CC or send an email to license@sib-sib.ch).

CC -----

CC EMBL: AF018635; AAB70306.1; -

CC InterPro: IPR002311; trna_synth_2f.

CC Pfam: PF02092; trna_synth_2f.1.

CC PROSITE: PS50861; AA_TRNA_LIGASE_IL_GLYAB; 1.

CC AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.

CC NON_TER

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

CC KW AMINOACYL-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;

CC Complete proteome.

Db 563 MKAVSHRTLDAAALAAANKRVSNILAKSDEVLSDRVNASTLIKEPEIKLAWGVVLRD 622
 QY 60 KIAAFAEGNFRRLTSLASAKPOVDAFPDGVWMAEDPAKONRLNLNLAPOMNAYA 119
 Db 623 KLEPYTEGRYDALVELAELEPVDAFPDGVWMAEDKELRINRLIMLEKRLFLRYA 682
 QY 120 DIALL 124
 Db 683 DISLL 687

RESULT 6

SYGB_COXBU STANDARD; PRT; 284 AA.
 ID SYGB_COXBU
 AC P45651;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase beta chain) (GlyRS) (Fragment).
 GN GLYS.
 OS Coxiella burnetii.
 OC Bacteria; Proteobacteria; gamma subdivision; Legionellaceae group;
 OC Coxiella group; Coxiella.
 OX NCBI_TaxID=777;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Nine Mile phase I / Bratislava;
 RX MEDLINE=94350801; PubMed=8071197;
 RA Susan M., Chen S.Y., Thompson H.A., Hoover T.A., Hill A.,
 RA Williams J.C.;
 RT "Cloning and characterization of an autonomous replication sequence from Coxiella burnetii".
 RL J. Bacteriol. 176:5233-5243(1994).
 RN [2]
 RP CONCEPTUAL TRANSLATION.
 RA Bairoch A.;
 CC Unpublished observations (AUG-1995).
 CC -1- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate + glycyl-tRNA(Gly).
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains.
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -1- CAUTION: THIS IS A CONCEPTUAL TRANSLATION; THREE FRAMESHIFTS WERE INTRODUCED TO MAXIMIZE THE SIMILARITY WITH ORTHOLOGS IN OTHER BACTERIA.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: U10529; AAA56913.1; ALT_FRAME.
 CC InterPro: IPR002311; ATRNA_ligaseI.
 CC InterPro: IPR002311; tRNA_synth_2f.
 CC Pfam: PF02092; tRNA_synth_2f.1
 CC DR PROSITE: PS50861; AA_TRNA_Ligase_I1_GLYAB.1.
 CC K04 Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding.
 CC FT NON_TER
 CC SEQUENCE 284 AA; 32048 MW; 3B76C8421076FF11 CRC64;

Query Match 39.8%; Score 238; DB 1; Length 284;
 Best Local Similarity 44.4%; Pred. No. 5.3e-13;
 Matches 55; Conservative 21; Mismatches 48; Indels 0; Gaps 0;

QY 1 LOAVAVFKOLPEAAALAAANKRVONILKKAADAL-GEVNESILQODEEKALYAAAGGIQPK 60
 Db 160 IKAVOHFQTLPEADALAAANKRVSNILKKADELKSIDHSLPFDSDAHLADQLKRAEL 219
 QY 61 IAAVAEGNFRRLTSLASAKPOVDAFPDGVWMAEDPAKONRLNLNLAPOMNAYA 120

Db 220 VNNLYKRDYTKALSELASLKEPDIIDFDVWMAEDKEREKRNRLALLSSLOQFSGIAD 279
 QY 121 IALL 124
 Db 280 ISLL 283

RESULT 7

SYGB_PASMU STANDARD; PRT; 689 AA.
 ID SYGB_PASMU
 AC P57905;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase beta chain) (GlyRS).
 GN GLYS OR PM1102
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=Pm70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida Pm70".
 RL Proc Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 CC -1- CATALYTIC ACTIVITY: ATP + glycine + tRNA(Gly) = AMP + diphosphate + glycyl-tRNA(Gly).
 CC -1- SUBUNIT: Tetramer of two alpha and two beta chains (By similarity).
 CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -1- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
 CC
 CC EMBL: AE006151; AAK03186.1; -
 CC InterPro: IPR002311; tRNA_synth_2f.
 CC Pfam: PF02092; tRNA_synth_2f.1.
 CC DR PRINTS: PR01045; TRNASYNTHGB.
 CC DR TIGRFAMS: TIGR00211; GLYS.1.
 CC DR PROSITE: PS50861; AA_TRNA_Ligase_I1_GLYAB.1.
 CC KW Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding; Complete proteome.
 CC SEQUENCE 689 AA; 76019 MW; D67A980B5143B21E CRC64;

Query Match 38.7%; Score 231.5; DB 1; Length 689;
 Best Local Similarity 43.2%; Pred. No. 4.8e-12;
 Matches 54; Conservative 23; Mismatches 47; Indels 1; Gaps 1;

QY 1 LOAVAVFKOLPEAAALAAANKRVONILKKAADAL-GEVNESILQODEEKALYAAAGGIQPK 59
 Db 564 VRAVSIFRLEAEALAAANKRVSNILAVVEGELPANIDITTLCAEAKEVLAEOVYALQA 623
 QY 60 KIAAFAEGNFRRLTSLASAKPOVDAFPDGVWMAEDPAKONRLNLNLAPOMNAYA 119
 Db 624 ELAPLFAKGYVALDRLAALREPVDTFFDVMVNAENPOLRONRLAILNNLEFLQYA 683
 QY 120 DIALL 124
 Db 684 DISLL 688

RESULT 8

SYGB_XYLFRA

DR EMBL; 299117; CAB14468.1; -.
 DR Subtilist; BG11658; glys.
 DR InterPro; IPR002311; tRNA_synth_2f.
 DR Pfam; PF02092; tRNA_synth_2f; 1.
 DR PRINTS; PRO1045; TRNASYNTHGB.
 DR TIGRFAMS; TIGR00211; glys; 1.
 DR PROSITE; PS50861; AA_tRNA_LIGASE_II_GLYAB; 1.
 DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 SQ SEQUENCE 679 AA; 76232 MW; ACC96723A85922B CRC64;

Query Match 23.1%; Score 138; DB 1; Length 679;
 Best Local Similarity 28.8%; Pred. No. 0.0027;
 Matches 34; Conservative 31; Mismatches 45; Indels 8; Gaps 2;

0Y 7 PROPEAALAAANKRVONLKKADAALGEVNESLLOODEEKLVAAGCLOPKIAAAY 66
 DB 567 FETMEALG-----RVISISKK--GVREDIDPDLEFENEYKALPDVAYTAENIQENFS 618
 0Y 67 EGNFTALSELASVPRVDVAFDGVVMAEDAAYKONRLNLRLAEQNNAYADIAL 124
 DB 619 KNDYEALALASLAKPEPIDAYFDHMTVADNESLKANRLAQNVSILADEIKSFANMAL 676

RESULT 12
 SYGB_AOUAE STANDARD; PRT; 664 AA.
 AC 067898;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase
 beta chain) (GLYS).
 GN GLYS OR AQ.2141.
 OS Aquifex aeolicus.
 OC Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
 OC Aquifex.
 OX NCBI_TaxID=63363;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-VF5.
 RX MEDLINE-9819666; PubMed-9537320;
 RA Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
 RA Graham D.E., Overbeek R., Sneed M.A., Keller M., Aujay M., Huber R.,
 RA Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
 RT "The complete genome of the hyperthermophilic bacterium Aquifex
 aeolicus".
 RL Nature 392:353-358(1998).
 CC -i- CATALYTIC ACTIVITY: ATP + glycine + tRNA(gly) = AMP + diphosphate
 CC + glycyl-tRNA(gly).
 CC -i- SUBUNIT: Tetramer of two alpha and two beta chains
 CC (by similarity).
 CC -i- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -i- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AB000775; AAC07870.1; -.
 DR InterPro; IPR002311; tRNA_synth_2f.
 DR Pfam; PF02092; tRNA_synth_2f; 1.
 DR PRINTS; PRO1045; TRNASYNTHGB.
 DR TIGRFAMS; TIGR00211; glys; 1.
 DR PROSITE; PS50861; AA_tRNA_LIGASE_II_GLYAB; 1.
 DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 SQ SEQUENCE 664 AA; 77523 MW; 7FEF7A8F552E3DB CRC64;

Query Match 22.9%; Score 137; DB 1; Length 664;
 Best Local Similarity 33.6%; Pred. No. 0.00032;
 Matches 37; Conservative 18; Mismatches 45; Indels 10; Gaps 3;

0Y 11 PEAALAAANKRVONLKKADAALGEVNESLLOODEEKLVAAGCLOPKIAAAYAE 70
 DB 556 PEPFSVVEAKKRVIRIIRPK-DWKNEYEVDEKLLSEERELYOKLTREFENK-----EL 606
 0Y 71 FTALSELASVPRVDVAFDGVVMAEDAAYKONRLNLRLAEQNNAYAD 120
 DB 607 KSP-L-ELPLKEYIDKFVDNVKVAEDEKIRNNRISLKRVENLFRFTFG 655

RESULT 13
 SYGB_BACHD STANDARD; PRT; 693 AA.
 AC 09KD48;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14) (Glycine--tRNA ligase
 beta chain) (GLYS).
 GN GLYS OR BH1371.
 OS Bacillus halodurans.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=86665;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-C-125 / JCM 9153;
 RX MEDLINE-20512582; PubMed-11058132;
 RA Takami H., Nakasone K., Takaki Y., Maeno G., Sasaki R., Masui N.,
 RA Fuji F., Hirama C., Nakamura Y., Ogasawara N., Kuhnara S.,
 RA Horikoshi K.;
 RT "Complete genome sequence of the alkaliphilic bacterium Bacillus
 halodurans and genomic sequence comparison with Bacillus subtilis".
 RL Nucleic Acids Res. 28:4317-4331(2000).
 CC -i- CATALYTIC ACTIVITY: ATP + glycine + tRNA(gly) = AMP + diphosphate
 CC + glycyl-tRNA(gly).
 CC -i- SUBUNIT: Tetramer of two alpha and two beta chains
 CC (by similarity).
 CC -i- SUBCELLULAR LOCATION: Cytoplasmic.
 CC -i- SIMILARITY: BELONGS TO CLASS-II AMINOACYL-TRNA SYNTHETASE FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL; AP001511; BAB05090.1; -.
 DR InterPro; IPR002311; tRNA_synth_2f.
 DR Pfam; PF02092; tRNA_synth_2f; 1.
 DR PRINTS; PRO1045; TRNASYNTHGB.
 DR TIGRFAMS; TIGR00211; glys; 1.
 DR PROSITE; PS50861; AA_tRNA_LIGASE_II_GLYAB; 1.
 DR Aminoacyl-tRNA synthetase; Protein biosynthesis; Ligase; ATP-binding;
 KW Complete proteome.
 SQ SEQUENCE 693 AA; 77833 MW; 30F9919712105754 CRC64;

Query Match 22.9%; Score 137; DB 1; Length 693;
 Best Local Similarity 28.2%; Pred. No. 0.00034;
 Matches 35; Conservative 26; Mismatches 35; Indels 10; Gaps 2;

0Y 6 VERQ-----LPEAALAAANKRVONLKKADAALGEVNESLLOODEEKLVAAG 56
 DB 564 VPKKAKILVSKVNPDEFKELVGLSRVTNIAKAEKNA-IPNDFKEKEEVLVEAYVQ 622
 0Y 57 LOPKIAAAYAEENFTALSELASVPRVDVAFDGVVMAEDAAYKONRLNLRLAEQNN 116
 DB 623 TRDLVQALASGDSVAAYAALEDTTIEPIHGYPEHVWVWVDEQVIRENRLAHAFAGVIG 682

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 10.8332 Seconds
(without alignments)
2398.732 Million cell updates/sec

Title: US-10-007-267-2

Perfect score: 598

Sequence: 1 LQAVAVFKQLPEALAAAN.....LTNRLAEQNNNAVADIALIGE 126

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp.archaea:*
2: sp.bacteria:*
3: sp.fungi:*
4: sp.human:*
5: sp.invertebrate:*
6: sp.mammal:*
7: sp.mhc:*
8: sp.organelle:*
9: sp.phage:*
10: sp.plant:*
11: sp.rodent:*
12: sp.virus:*
13: sp.vertibrate:*
14: sp.unclassified:*
15: sp.virus:*
16: sp.bacteriap:*
17: sp.archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	575	96.2	687	16	Q9JX05
2	572	95.7	687	16	Q9JW67
3	408	68.2	87	2	Q51114
4	313	52.3	697	16	Q8Y213
5	256.5	42.9	684	16	Q917B8
6	252.5	42.2	689	16	Q8XDN7
7	241.5	40.4	689	16	Q8Z9W7
8	229.5	38.4	689	16	Q8ZL96
9	229.5	38.4	689	16	Q8Z2B3
10	223	37.3	199	2	Q9F730
11	193	32.3	40	2	Q932S6
12	189	31.6	717	16	Q8UHN6
13	174	29.1	718	16	Q986B5
14	173	28.9	780	16	Q8YFJ1
15	164	27.4	673	16	Q9CGK0
16	160	26.8	678	16	Q97PW6

17	156	26.1	674	16	Q9ZG11	Q9ZG11 rickettsia
18	154	25.8	679	16	Q99Y14	Q99Y14 streptococc
19	154	25.8	721	16	Q92RL2	Q92RL2 rhizobium m
20	140.5	23.5	686	16	Q8RH44	Q8RH44 fusobacteri
21	138	23.1	692	16	Q9A8T1	Q9A8T1 caulobacter
22	124.5	20.8	716	16	Q8YPT0	Q8YPT0 anabaena sp
23	122	20.4	1067	10	Q23150	Q23150 arabislopsis
24	121	20.2	664	16	Q9PN63	Q9PN63 campylobact
25	119	19.9	1067	10	Q9SU73	Q9SU73 arabidopsis
26	103.5	17.3	688	16	Q8Y734	Q8Y734 listeria mo
27	92.5	15.5	688	16	Q92BQ2	Q92BQ2 listeria in
28	91	15.2	523	16	Q8YPF4	Q8YPF4 anabaena sp
29	91	15.2	1512	13	Q9DE36	Q9DE36 brachydanio
30	88	14.7	417	16	Q92LJ6	Q92LJ6 rhizobium m
31	88	14.7	1521	4	Q95710	Q95710 homo sapien
32	88	14.7	1525	4	Q9Y507	Q9Y507 homo sapien
33	88	14.7	1529	4	Q94813	Q94813 homo sapien
34	87	14.5	796	11	Q9WVC1	Q9WVC1 rattus norv
35	86.5	14.5	1010	16	Q8YK94	Q8YK94 anabaena sp
36	86	14.4	1025	11	Q92166	Q92166 mus musculu
37	86	14.4	1521	11	Q9R1B9	Q9R1B9 mus musculu
38	85.5	14.3	869	5	Q9NDA9	Q9NDA9 anisakis sl
39	85	14.2	1530	13	Q90WZ3	Q90WZ3 xenopus lae
40	84.5	14.1	423	17	Q9HQT5	Q9HQT5 halobacteri
41	84	14.0	522	5	Q18266	Q18266 caenorhabdi
42	84	14.0	804	1	P71416	P71416 halobacteri
43	83	13.9	449	2	Q91UV1	Q91UV1 escherichia
44	83	13.9	599	16	Q9KOC6	Q9KOC6 vibrio chol
45	82.5	13.8	259	4	Q9NXN5	Q9NXN5 homo sapien

ALIGNMENTS

RESULT 1
Q9JX05 PRELIMINARY: PRT: 687 AA.
ID Q9JX05
AC Q9JX05
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
DE Glycyl-tRNA synthetase, beta chain.
GN NMB1930.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; Beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
[1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / SEROGROUP B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tetelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Clecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Yamatelevan J.,
RA Gill J., Scarlato V., Masignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rapunoli R., Venter J.C.,
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
MC58."
RL Science 287:1809-1815 (2000).
DR EMBL: AE005541; AAF42259.1; -.
DR TIGR: NMB1930.
DR InterPro: IPR002311; tRNA_synth_2f.
DR Pfam: PF02092; tRNA_synth_2f; 1.
DR PRINTS: PR01045; TRNASYNTHCB.
DR TIGRFAMs: TIGR00211; g1ys; 1.
KW Aminoacyl-tRNA synthetase; Complete proteome.
SQ SEQUENCE 687 AA; 74573 MW; 616BCBDD76A3D4FB CRC64;

Query Match 96.2%; Score 575; DB 16; Length 687;
Best Local Similarity 96.0%; Pred. No. 5.6e-39;
Matches 121; Conservative 2; Mismatches 3; Indels 0; Gaps 0;


```

OY 1 LOAAVAFKOLPEAAALAAANKRVONLLKKADALGEVNESILLOODEKALYAAAGLOPK 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 562 LOAAVAFKOLPEAAALAAANKRVONLLKKADALGEVNESILLOODEKALFAAAGLOPK 621
OY 61 IAAVAEGNFRTALSELASVYPVDADFPGVYMAEDAAYKQNRNLNLRLAEQNNAVAD 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 622 IAAVAEGNFRTALSELASVYPVDADFPGVYMAEDAAYKQNRNLNLRLAEQNNAVAD 681
OY 121 IALIGE 126
    ||||| |||||
DB 682 IALIGE 687

RESULT 2
O9JW67 PRELIMINARY; PRT: 687 AA.
ID O9JW67
AC O9JW67;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Putative glycyl-tRNA synthetase beta subunit (EC 6.1.1.14).
GN GLYS OR NMA0523.
OS Neisseria meningitidis (serogroup A).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=65699;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
RX MEDLINE=2022256; PubMed=10761919;
RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
RA Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holroyd S.,
RA Jagels R., Leather S., Moule S., Mungall K., Quail M.A.,
RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
RA Whitehead S., Spratt B.G., Barrell B.G.;
RT "Complete DNA sequence of a serogroup A strain of Neisseria
RT meningitidis 22491."
RL Nature 404:502-506(2000).
DR EMBL: AL162753; CAB83815.1; -.
DR InterPro: IPR002311; tRNA_synt_2f.
DR Pfam: PF02092; tRNA_synt_2f.1.
DR PRINTS: PR01045; TRNASYNTHGB.
DR TIGRFAMs: TIGR00211; glys.1.
KW Aminoacyl-tRNA synthetase; Ligase; Complete proteome.
SQ SEQUENCE 687 AA; 74650 MW; 8C17811D5C93EBE3 CRC64;

Query Match 95.7%; Score 572; DB 16; Length 687;
Best Local Similarity 96.0%; Pred. No. 9.9e-39;
Matches 121; Conservative 2; Mismatches 3; Indels 0; Gaps 0;

OY 1 LOAAVAFKOLPEAAALAAANKRVONLLKKADALGEVNESILLOODEKALYAAAGLOPK 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 562 LOAAVAFKOLPEAAALAAANKRVONLLKKADALGEVNESILLOODEKALFAAAGLOPK 621
OY 61 IAAVAEGNFRTALSELASVYPVDADFPGVYMAEDAAYKQNRNLNLRLAEQNNAVAD 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 622 IAAVAEGNFRTALSELASVYPVDADFPGVYMAEDAAYKQNRNLNLRLAEQNNAVAD 681
OY 121 IALIGE 126
    ||||| |||||
DB 682 IALIGE 687

RESULT 3
O51114 PRELIMINARY; PRT: 87 AA.
ID O51114
AC O51114;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE glycyl tRNA synthetase (fragment).
GN GLYS.

```

```

OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58;
RX MEDLINE=96414473; PubMed=8817494;
RA Jennings M.P., Hood D., Peak I.R.A., Virji M., Moxon E.R.;
RT "Molecular analysis of a locus for the biosynthesis and phase-variable
RT expression of the lacto-N-neotetraose terminal lipopolysaccharide
RT structure in Neisseria meningitidis."
RL Mol. Microbiol. 18:729-740(1995).
RN [2]
RP SEQUENCE OF 12-87 FROM N.A.
RC STRAIN=126E;
RA Jennings M.P., Der Ley P., Poolman J.T., Moxon E.R.;
RL Submitted (MAR-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: U25839; AAC44083.1; -.
DR EMBL: U65788; AAB48383.1; -.
KW Aminoacyl-tRNA synthetase.
FT NON_TER 1 1
SQ SEQUENCE 87 AA; 9242 MW; 22990BBD6FA2C8E0 CRC64;

Query Match 68.2%; Score 408; DB 2; Length 87;
Best Local Similarity 97.7%; Pred. No. 2.2e-26;
Matches 85; Conservative 2; Mismatches 0; Indels 0; Gaps 0;

OY 40 SLLOODEEKALYAAAGLOPKIAAIVAEQNFRTALSELASVYPVDADFPGVYMAEDA 99
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 1 SLLOODEEKALFAAAGLOPKIAAIVAEQNFRTALSELASVYPVDADFPGVYMAEDA 60
OY 100 VKONRLNLRLAEQNNAVADIALIGE 126
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
DB 61 VKONRLNLRLAEQNNAVADIALIGE 87

RESULT 4
O8Y213 PRELIMINARY; PRT: 697 AA.
ID O8Y213
AC O8Y213;
DT 01-MAR-2002 (TREMblrel. 20, Created)
DT 01-MAR-2002 (TREMblrel. 20, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE Probable glycyl-tRNA synthetase beta chain protein (EC 6.1.1.14).
GN GLYS OR RSC0524 OR RSC0494.
OS Ralstonia solanacearum (Pseudomonas solanacearum).
OC Bacteria; Proteobacteria; beta subdivision; Ralstonia group;
OC Ralstonia.
OX NCBI_TaxID=305;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=GM11000;
RX MEDLINE=21681879; PubMed=11823852;
RA Salanoubat M., Genin S., Artiguenave F., Gouzy J., Mangenot S.,
RA Arlet M., Billault A., Brottier P., Camus J.C., Cattoilco L.,
RA Chandler M., Choise N., Claudel-Renard C., Cunac S., Demange N.,
RA Gaspin C., Layte W., Moisan A., Robert C., Saurin W., Schlex T.,
RA Sigulier P., Thebaud P., Whalen M., Wincker P., Levy M.,
RA Weissenbach J., Boucher C.A.;
RT "Genome sequence of the plant pathogen Ralstonia solanacearum."
RL Nature 415:497-502(2002).
DR EMBL: AL646059; CAD14052.1; -.
DR InterPro: IPR002311; tRNA_synt_2f.
DR Pfam: PF02092; tRNA_synt_2f.1.
DR PRINTS: PR01045; TRNASYNTHGB.
DR TIGRFAMs: TIGR00211; glys.1.
KW Ligase; Complete proteome.
SQ SEQUENCE 697 AA; 75257 MW; 15A133DAD177EBDE CRC64;

Query Match 52.3%; Score 313; DB 16; Length 697;
Best Local Similarity 53.2%; Pred. No. 1.5e-17;
Matches 66; Conservative 19; Mismatches 39; Indels 0; Gaps 0;

```


OY 1 LGAAVAFKQLPEAAALAAANKRVONILKKRADAALGEVNESLLQODEKALVYAAOGLOPK 60
 DB 572 LEAVAFALPQAEALAAANKRITITLTKKTDITIGSVQQLREDEKRALHQAVASEPH 631
 OY 61 IAAVAEENFRTALSELASVKPOVDAFEDGVVMAEDAAVKONRLNLRLAEQNNAVAD 120
 DB 632 VHAFAARGDFTTALKTALSLAREAVDSFFDGVVYVMAEDDTLRDNRLLALGELGLMNRVAD 691
 OY 121 IALL 124
 DB 692 ISKL 695

RESULT 5

OY17B8 PRELIMINARY: PRT: 684 AA.
 AC 0917B8.
 DT 01-MAR-2001 (TREMBLrel. 16, Created)
 DT 01-MAR-2001 (TREMBLrel. 16, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE GLYCYL-tRNA synthetase beta chain.
 GN GLYS OR PA0008.
 OS Pseudomonas aeruginosa.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=287;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-ATCC 15692 / PA01;
 RX MEDLINE=20437337; PubMed=10984043;
 RA Hickey C.K., Pham X.-Q.T., Erwin A.L., Mizoguchi S.D., Warren P.,
 RA Stover C.K., Brinkman F.S.L., Hufnagle W.O., Kowalik D.J., Lagrou M.,
 RA Garner R.L., Goltzy L., Tolentino E., Westbrook-Medman S., Yuan Y.,
 RA Brody L.L., Coulter S.N., Folger K.R., Kas A., Laidig K., Lim R.M.,
 RA Smith K.A., Spencer D.H., Wong G.K.-S., Wu Z., Paulsen I.T.,
 RA Reiter J., Sater M.H., Hancock R.E.W., Lory S., Olson M.V.,
 RA "Complete genome sequence of Pseudomonas aeruginosa PA01, an
 RT opportunistic pathogen."
 RT Nature 406:959-964(2000).
 RL EMBL: AE004440; AAC03398.1; -
 DR InterPro: IPR002311; tRNA_synth_2f.
 DR Pfam: PF02092; tRNA_synth_2f; 1.
 DR PRINTS: PR01045; TRNASYNTHGB.
 DR TIGRFAMS: TIGR00211; GLYS; 1.
 KM Aminoacyl-tRNA synthetase; Complete proteome.
 SQ SEQUENCE 684 AA; 73973 MW; 74144075010F0BB6 CRC64;

Query Match 42.9%; Score 256.5; DB 16; Length 684;
 Best Local Similarity 48.4%; Pred. No. 6.2e-13;
 Matches 61; Conservative 20; Mismatches 44; Indels 1; Gaps 1;

OY 1 LGAAVAFKQLPEAAALAAANKRVONILKKRADAALGEVNESLLQODEKALVYAAOGLOPK 59
 DB 559 VOAVDAFRLPPEALAAANKRVSNILAKSDEVPNPVADSLVEAKALGSAAVANAES 618
 OY 60 KTAANAEGNFRTALSELASVKPOVDAFEDGVVMAEDAAVKONRLNLRLAEQNNAVAD 119
 DB 619 EVAFLAARDYRAALARLALALREPDVTFADYVAVVNDAAVANKRYALALAKLGSFLGVA 678
 OY 120 DIALG 125
 DB 679 DISLIG 684

RESULT 6

OY8XDN7 PRELIMINARY: PRT: 689 AA.
 AC 08XDN7.
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE Glycine tRNA synthetase, beta subunit.
 GN GLYS OR 24983 OR ECS4442.

OS Escherichia coli O157:H7.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OX NCBI_TaxID=83334;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / EDL933 / ATCC 700927;
 RX MEDLINE=21074935; PubMed=11206551;
 RA Perna N.T., Plunkett G., III, Burland V., Mau B., Glasner J.D.,
 RA Rose D.J., Mayhew G.F., Evans P.S., Gregor J., Kirkpatrick H.A.,
 RA Postel G., Hackett J., Link S., Boutin A., Shao Y., Miller L.,
 RA Grobeck E.J., Davis N.W., Lim A., Dimalanta E.T., Pocanousis K.,
 RA Apodaca J., Anantharaman T.S., Lin J., Yen G., Schwartz D.C.,
 RA Welch R.A., Blattner F.R.;
 RT "Genome sequence of enterohaemorrhagic Escherichia coli O157:H7."
 RT Nature 409:529-533(2001).
 RL Nature 409:529-533(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-O157:H7 / RIMD 0509952;
 RX MEDLINE=21156231; PubMed=11258796;
 RA Hayashi T., Makino K., Ohnishi M., Kurokawa K., Ishii K., Yokoyama K.,
 RA Han C.-G., Ohnishi E., Nakayama K., Murata T., Tanaka M., Tobe T.,
 RA Iida T., Takami H., Honda T., Sasakawa C., Ogasawara N., Yasunaga T.,
 RA Kuhara S., Shiba T., Hattori M., Shinagawa H.;
 RT "Complete genome sequence of enterohaemorrhagic Escherichia coli
 RT O157:H7 and genomic comparison with a laboratory strain K-12.";
 RL DNA Res. 8:11-22(2001).
 DR EMBL: AE005582; AAG58707.1; -
 DR EMBL: AP002565; BAB37865.1; -
 DR InterPro: IPR002311; tRNA_synth_2f.
 DR Pfam: PF02092; tRNA_synth_2f; 1.
 DR PRINTS: PR01045; TRNASYNTHGB.
 DR TIGRFAMS: TIGR00211; GLYS; 1.
 KM Aminoacyl-tRNA synthetase; Complete proteome.
 SQ SEQUENCE 689 AA; 76813 MW; 3F9B939FEA3168CD CRC64;

Query Match 42.2%; Score 252.5; DB 16; Length 689;
 Best Local Similarity 48.0%; Pred. No. 1.3e-12;
 Matches 60; Conservative 21; Mismatches 43; Indels 1; Gaps 1;

OY 1 LGAAVAFKQLPEAAALAAANKRVONILKKRADAALGEVNESLLQODEKALVYAAOGLOPK 59
 DB 564 MKAVSHFTLEAAALAAANKRVSNILAKSDEVLSDRVNAVSTLKEPEIKLMAQVVLVD 623
 OY 60 KTAANAEGNFRTALSELASVKPOVDAFEDGVVMAEDAAVKONRLNLRLAEQNNAVAD 119
 DB 624 KLEPYFAEGRYDALVELALREPDVTFADYVAVVNDDELRLNRLTMEKLEFLRYVA 683
 OY 120 DIALG 124
 DB 684 DISL 688

RESULT 7

OY8Z9W7 PRELIMINARY: PRT: 689 AA.
 AC 08Z9W7.
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, Last annotation update)
 DE GLYCYL-tRNA synthetase beta subunit (EC 6.1.1.14).
 GN GLYS OR YPO4071.
 OS Yersinia pestis.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Yersinia.
 OX NCBI_TaxID=632;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-CO-92 / BIOVAR ORIENTALIS;
 RX MEDLINE=21470413; PubMed=11586360;
 RA Parhill J., Wren B.W., Thomson N.R., Tibball R.W., Holden M.T.G.,
 RA Prentice M.B., Sebahia M., James K.D., Churcher C., Mungall K.L.,
 RA Baker S., Basham D., Bentley S.D., Brooks K., Cerdeno-Tarraga A.M.,

RA Chillingworth T., Cronin A., Davies R.M., Davis P., Dougan G.,
 RA Feltwell T., Hamlin N., Holroyd S., Jagels K., Rathliff A.V.,
 RA Leather S., Moule S., Oyston P.C.F., Quail M., Rutherford K.,
 RA Simmonds M., Skelton J., Stevens K., Whitehead S., Barrell B.G.,
 RT "genome sequence of Yersinia pestis, the causative agent of plague."
 RL Nature 413:523-527(2001).
 DR EMBL: AJ414160; CAC93524.1; -
 DR InterPro: IPR002311; trna_synth_2f.
 DR Pfam: PF02092; trna_synth_2f; 1.
 DR PRINTS: PRO1045; TRNASYNTHGB.
 DR TIGRFAMs: TIGR00211; glys; 1.
 KW Antinocyl-tRNA synthetase; Ligase; Complete proteome.
 SQ SEQUENCE 689 AA; 76204 MW; 715AF2AEED659397 CRC64;

Query Match 40.4%; Score 241.5; DB 16; Length 689;
 Best Local Similarity 46.4%; Pred. No. 1,1e-11;
 Matches 58; Conservative 23; Mismatches 43; Indels 1; Gaps 1;

OY 1 LOAAVAFKQLEPAALAAANKRVONLKKADAALGE-VNESLLOODEKALYAAAGIOP 59
 DB 564 VKAVYFRTLLAALAAANKRVSNILAKSTDTLDHVASITLKPALKTTLHLVLRD 623
 OY 60 KIAAFAEGNFRFALSELASVKPOVDAFDGVVMAEDAAVKONRLNRLAEQNNAVA 119
 DB 624 QLEPFFAAGQIKELALVELALRETYDEFFESVMAEDAAVNRLLTILSKIRELFQYA 683
 OY 120 DIALL 124
 DB 684 DISL 688

RESULT 8

O82L96 PRELIMINARY; PRT; 689 AA.
 AC O82L96;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Glycine tRNA synthetase, beta subunit (EC 6.1.1.14).
 GN GLYS OR STM3655.
 OS Salmonella typhimurium.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=602;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=LT2 / SGSC1412 / ATCC 700720;
 RX MEDLINE=21534948; PubMed=11677609;
 RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
 RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
 RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
 RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
 RA Waterston R., Wilson R.K.;
 RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
 LT2.";
 RL Nature 413:852-856(2001).
 DR EMBL: AB008869; AAL2514.1; -
 DR InterPro: IPR002311; trna_synth_2f.
 DR Pfam: PF02092; trna_synth_2f; 1.
 DR PRINTS: PRO1045; TRNASYNTHGB.
 DR TIGRFAMs: TIGR00211; glys; 1.
 KW Aminoacyl-tRNA synthetase; Ligase; Complete proteome.
 SQ SEQUENCE 689 AA; 76452 MW; 5734C31C5223123F CRC64;

Query Match 38.4%; Score 229.5; DB 16; Length 689;
 Best Local Similarity 44.8%; Pred. No. 1e-10; Indels 1; Gaps 1;
 Matches 56; Conservative 26; Mismatches 42; Indels 1; Gaps 1;

OY 1 LOAAVAFKQLEPAALAAANKRVONLKKADAALGE-VNESLLOODEKALYAAAGIOP 59
 DB 564 MKAVSHFRTLEASALAAANKRVSNILAKATEPLNDIVHASVLKEAEITLARHLVLRD 623
 OY 60 KIAAFAEGNFRFALSELASVKPOVDAFDGVVMAEDAAVKONRLNRLAEQNNAVA 119

DB 624 KLOPYFADRGQELIELALRAPVDEFFENVVMAEKDIRINRLTILSKIRELFQYA 683
 OY 120 DIALL 124
 DB 684 DISL 688

RESULT 9

O82ZB3 PRELIMINARY; PRT; 689 AA.
 AC O82ZB3;
 DT 01-MAR-2002 (TREMBlrel. 20, Created)
 DT 01-MAR-2002 (TREMBlrel. 20, Last sequence update)
 DT 01-JUN-2002 (TREMBlrel. 21, Last annotation update)
 DE Glycine tRNA synthetase, beta subunit (EC 6.1.1.14).
 GN STY4144.
 OS Salmonella typhi.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Salmonella.
 OX NCBI_TaxID=601;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=C718;
 RX MEDLINE=21534947; PubMed=11677608;
 RA Parkhill J., Dougan G., James K.D., Thomson N.R., Pickard D., Main J.,
 RA Churcher C., Mungall K.L., Bentley S.D., Holden M.T.G., Seabaugh M.,
 RA Baker S., Basham D., Brooks K., Chillingworth T., Connerton P.,
 RA Cronin A., Davis P., Davies R.M., Dowd L., White N., Farrar J.,
 RA Feltwell T., Hamlin N., Haque A., Hien T.T., Holroyd S., Jagels K.,
 RA Krogh A., Larsen T.S., Leather S., Moule S., O'Gaora P., Parry C.,
 RA Quail M., Rutherford K., Simmonds M., Skelton J., Stevens K.,
 RA Whitehead S., Barrell B.G.;
 RT "Complete genome sequence of a multiple drug resistant Salmonella
 enterica serovar Typhi CT18.";
 RL Nature 413:848-852(2001).
 DR EMBL: AL627281; CAD07973.1; -
 DR InterPro: IPR002311; trna_synth_2f.
 DR Pfam: PF02092; trna_synth_2f; 1.
 DR PRINTS: PRO1045; TRNASYNTHGB.
 DR TIGRFAMs: TIGR00211; glys; 1.
 KW Aminoacyl-tRNA synthetase; Ligase; Complete proteome.
 SQ SEQUENCE 689 AA; 76645 MW; A93C239351028D9 CRC64;

Query Match 38.4%; Score 229.5; DB 16; Length 689;
 Best Local Similarity 44.8%; Pred. No. 1e-10; Indels 1; Gaps 1;
 Matches 56; Conservative 26; Mismatches 42; Indels 1; Gaps 1;

OY 1 LOAAVAFKQLEPAALAAANKRVONLKKADAALGE-VNESLLOODEKALYAAAGIOP 59
 DB 564 MKAVSHFRTLEASALAAANKRVSNILAKATEPLNDIVHASVLKEAEITLARHLVLRD 623
 OY 60 KIAAFAEGNFRFALSELASVKPOVDAFDGVVMAEDAAVKONRLNRLAEQNNAVA 119
 DB 624 KLOPYFADRGQELIELALRAPVDEFFENVVMAEKDIRINRLTILSKIRELFQYA 683
 OY 120 DIALL 124
 DB 684 DISL 688

RESULT 10
 O9F730 PRELIMINARY; PRT; 199 AA.
 ID O9F730;
 AC O9F730;
 DT 01-MAR-2001 (TREMBlrel. 16, Created)
 DT 01-MAR-2001 (TREMBlrel. 16, Last sequence update)
 DT 01-JUN-2001 (TREMBlrel. 17, Last annotation update)
 DE Glys-like protein (Fragment).
 OS Pseudomonas fluorescens.
 OC Bacteria; Proteobacteria; gamma subdivision; Pseudomonadaceae;
 OC Pseudomonas.
 OX NCBI_TaxID=294;

RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-F113;
 RX MEDLINE=20476421; PubMed=11021923;
 RA Laue B.E., Jiang Y., Chhabra S.R., Jacob S., Stewart G.S.A.B.,
 RT "The biocontrol strain *Pseudomonas fluorescens* F113 produces the
 RT rhizobium small bacteriocin, N-(3-hydroxy-7-cis-
 RT tetradecenyl)homoserine lactone, via Hds, a putative novel N-
 RT acylhomoserine lactone synthase."
 RL Microbiology 146:2469-2480(2000).
 DR EMBL: AF286536; ANG30824.1; -;
 DR Interpro: IPR002311; tRNA_synth_2f.
 DR Pfam: PF02092; tRNA_synth_2f; 1.
 FT NON_TER 1
 SQ SEQUENCE 199 AA; 21900 MW; 464CD941AA21A7C1 CRC64;

Query Match 37.3%; Score 223; DB 2; Length 199;
 Best Local Similarity 40.8%; Pred. No. 7.9e-11;
 Matches 51; Conservative 26; Mismatches 48; Indels 0; Gaps 0;

QY 1 LOAAVAFKQLEPAALAAANKRVQNLTKADALGGEVNSLQQDEKALYAAAGLQPK 60
 Db 75 VQAVQAFKLEPAALAAANKRVSNLSKAEPSYKAVEARYFDNAEFNSAIQQAAYAS 134
 QY 61 IAAVAEGNFRALSELASVYPQVDAFFDGVWMAEDAAVKONRLNLRLAEQNNVAD 120
 Db 135 VQPLMEKROYAELARLALREPVDAAFEFAVWMAEDVEDYKKNRYLALRLSLFLNIAD 194
 QY 121 IALLG 125
 Db 195 ISVLG 199

RESULT 11
 Q932S6 PRELIMINARY; PRT; 40 AA.
 AC Q932S6;
 DT 01-DEC-2001 (TrEMBLrel. 19, Created)
 DT 01-DEC-2001 (TrEMBLrel. 19, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE GLYS (Fragment).
 GN GLYS.
 OS Neisseria meningitidis.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 NCBI_TaxID=487;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-M978, AND A1;
 RX MEDLINE=21467954; PubMed=11583844;
 RA Zhu P., Klutch M.J., Tsai C.-M.;
 RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
 RT Expression in Two 18-Immuno-type Strains of *Neisseria meningitidis*."
 RL FEMS Microbiol. Lett. 203:173-177(2001).
 DR EMBL: AF355193; AAL12837.1; -;
 DR EMBL: AF355194; AAL12843.1; -;
 FT NON_TER 1
 SQ SEQUENCE 40 AA; 4392 MW; E41CA1A42B65A187 CRC64;

Query Match 32.3%; Score 193; DB 2; Length 40;
 Best Local Similarity 100.0%; Pred. No. 3.3e-09;
 Matches 40; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 87 FPDGVWMAEDAAVKONRLNLRLAEQNNVADIALGE 126
 Db 1 FPDGVWMAEDAAVKONRLNLRLAEQNNVADIALGE 40

RESULT 12
 Q8UHN6 PRELIMINARY; PRT; 717 AA.
 AC Q8UHN6;
 DT 01-JUN-2002 (TrEMBLrel. 21, Created)

DT 01-JUN-2002 (TrEMBLrel. 21, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE GLYC1-CRNA synthetase, beta chain.
 GN GLYS OR ATU0644 OR AGR C.1144.
 OS Agrobacterium tumefaciens (strain C58 / ATCC 33970).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Rhizobium.
 NCBI_TaxID=176299;
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608550; PubMed=11743193;
 RA Wood D.W., Setubal J.C., Kaul R., Monks D.E., Kitajima J.P.,
 RA Okura Y.K., Zhou Y., Chen L., Wood G.E., Almeida N.F., Jr., Woo L.,
 RA Chen Y., Paulsen I.T., Eisen J.A., Karp P.D., Boyce D. Sr.,
 RA Chapman P., Clendenning J., Deatherage G., Gillet W., Grant C.,
 RA Kutayavin T., Levy R., Li M.-J., McClelland E., Palmeri A.,
 RA Raymond C., Rouse G., Saenphitachak C., Wu Z., Romero P., Gordon D.,
 RA Zhang S., Yoo H., Tao Y., Biddle P., Jung M., Krespan W., Perry M.,
 RA Gordon-Kamm B., Liao L., Kim S., Hendrick C., Zhao Z.-Y., Dolan M.,
 RA Chumley F., Tingey S.V., Tomb J.-F., Gordon M.P., Olson M.V.,
 RA Nester E.W.;
 RT "The genome of the natural genetic engineer *Agrobacterium tumefaciens*
 RT C58."
 RT Science 294:2317-2323(2001).
 RN [2]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=21608551; PubMed=11743194;
 RA Goodner B., Hinkle G., Gattung S., Miller N., Blanchard M.,
 RA Outullo B., Goldman B.S., Cao Y., Askenazi M., Halling C., Mullin L.,
 RA Houmlet K., Gordon J., Vaudin M., Iatchouk O., Epp A., Liu F.,
 RA Wollam C., Allinger M., Dougherty D., Scott C., Lappas C., Markelz B.,
 RA Flanagan C., Crowell C., Gursion J., Lomo C., Sear C., Strub G.,
 RA Cielo C., Slater S.;
 RT "Genome sequence of the plant pathogen and biotechnology agent
 RT Agrobacterium tumefaciens C58."
 RT Science 294:2323-2328(2001).
 DR EMBL: AE009032; AAL41660.1; -;
 DR EMBL: AE007998; AAK86451.1; -;
 KW Aminocycl-1-cRNA synthetase: Complete proteome.
 SQ SEQUENCE 717 AA; 78713 MW; 94F26877B8EF048 CRC64;

Query Match 31.6%; Score 189; DB 16; Length 717;
 Best Local Similarity 38.1%; Pred. No. 2.2e-07;
 Matches 48; Conservative 20; Mismatches 54; Indels 4; Gaps 1;

QY 1 LOAAVAFKQLEPAALAAANKRVQNLTKADALGGEVNSLQQDEKALYAAAGLQPK 60
 Db 588 VEAALTAFTTGEDGRNLLAGKARATQLLAAEEKGTVAADVSEELKLDKALYAAIKT 647
 QY 57 LOPKIAAVALGNGFRALSELASVYPQVDAFFDGVWMAEDAAVKONRLNLRLAEQNNVAD 116
 Db 648 ASADAAKAAVGEDEPRSMQALSTIRAPVDFEEDVLYNDEDAIRANRLALAIRATG 707
 QY 117 AVADIA 122
 Db 708 TVADFS 713

RESULT 13
 Q986B5 PRELIMINARY; PRT; 718 AA.
 AC Q986B5;
 DT 01-OCT-2001 (TrEMBLrel. 18, Created)
 DT 01-OCT-2001 (TrEMBLrel. 18, Last sequence update)
 DT 01-JUN-2002 (TrEMBLrel. 21, Last annotation update)
 DE GLYC1-CRNA synthetase, beta chain.
 GN MLR7435.
 OS Rhizobium loti (Mesorhizobium loti).
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Rhizobiaceae; Mesorhizobium.
 NCBI_TaxID=381;
 RN [1]
 RP SEQUENCE FROM N.A.

RC STRAIN=MAFE303099;
 RA MEDLINE=21082930; PubMed=11214968;
 RX Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
 RA Matsunabe A., Idesawa K., Ishikawa A., Kawashima K., Kimura T.,
 RA Kishida Y., Kiyokawa C., Kohara M., Matsunoto M., Matsuno A.,
 RA Mochizuki Y., Nakayama S., Nakazaki N., Shimo S., Sugimoto M.,
 RA Takeuchi C., Yamada M., Tabata S.;
 RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
 Mesorhizobium loti.";
 RT DNA Res. 7:331-338(2000).
 RL EMBL, AP003011; BAB53538.1; -
 DR InterPro: IPR002311; tRNA_synth_2f.
 DR Pfam: PF02092; tRNA_synth_2f; 1.
 DR PRINTS: PRO1045; TRNASYNTHGB.
 DR TIGRFAMS: TIGR00211; glys; 1.
 KW Aminocyl-tRNA synthetase; Complete proteome.
 SQ SEQUENCE 718 AA; 78610 MW; 66640296CFAC669 CRC64;

Query Match 29.1%; Score 174; DB 16; Length 718;
 Best Local Similarity 32.5%; Pred. No. 3.7e-06;
 Matches 41; Conservative 26; Mismatches 55; Indels 4; Gaps 1;

OY 1 LOAVAVFKQLEPAALAAANKRVQNL-----KKADALGEVNESLLOODEKALYAAAG 56
 Db 589 VEALGSFDTEDGKNLGTRAAANILAAEEKKTAQVAKTPEALPKENAEKSLFAAVNQ 648
 OY 57 LQPKIAAIVAEGNFRITLSELASVKKPOVDAFPDGVMAEDAFAVQNRNLINRLAEQMN 116
 Db 649 AEKQAGEAIQNDSESAHIALSALREPDSPFEGVLVNDDELEVANRALLTRIRATG 708
 OY 117 AVADIA 122
 Db 709 QVADFS 714

RESULT 14

OBYFJ1 PRELIMINARY; PRT; 780 AA.
 AC 08YFJ1;
 DT 01-MAR-2002 (TREMBLrel. 20, Created)
 DT 01-MAR-2002 (TREMBLrel. 20, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14).
 GN BMEI1529.
 OS Brucella melitensis.
 OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
 OC Brucellaceae; Brucella.
 OX NCBI_TaxID=29459;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=16M / ATCC 23456 / BIOTYPE 1;
 RX MEDLINE=20020109; PubMed=11756688;
 RA Delvecchio V.G., Kaputal V., Redkar R.J., Patra G., Mujer C., Los T.,
 RA Ivanova N., Anderson I., Bhatnagar A., Lykdis A., Resnik G.,
 RA Jablonski L., Larsen N., D'Souza M., Bernal A., Mazur M., Goldsman E.,
 RA Selkov E., Elzer P.H., Hagius S., O'Callaghan D., Detesoon J.-J.,
 RA Haselkorn R., Kyridides N., Overbeek R.;
 RT "The genome sequence of the facultative intracellular pathogen
 Brucella melitensis.";
 RT Proc. Natl. Acad. Sci. U.S.A. 99:443-448(2002).
 RL EMBL, AE009588; AAL52710.1; -
 DR InterPro: IPR002311; tRNA_synth_2f.
 DR Pfam: PF02092; tRNA_synth_2f; 1.
 DR TIGRFAMS: TIGR00211; glys; 1.
 KW Ligase; Complete proteome.
 SQ SEQUENCE 780 AA; 85795 MW; 27947D756B21FE80 CRC64;

Query Match 28.9%; Score 173; DB 16; Length 780;
 Best Local Similarity 35.7%; Pred. No. 4.9e-06;
 Matches 45; Conservative 22; Mismatches 55; Indels 4; Gaps 1;

OY 1 LOAVAVFKQLEPAALAAANKRVQNL-----KKADALGEVNESLLOODEKALYAAAG 56

Db 651 LEALIVFNEVDGKNLLAGAKRANILAAEEKKGTQVADVPALPRLREAEKALFADVTL 710
 OY 57 LQPKIAAIVAEGNFRITLSELASVKKPOVDAFPDGVMAEDAFAVQNRNLINRLAEQMN 116
 Db 711 ASREVGEAIQNDSESAHIALSALREPDSPFEGVLVNDDELEVANRALLTRIRATG 770
 OY 117 AVADIA 122
 Db 771 KVADFS 776

RESULT 15

O9CGKO PRELIMINARY; PRT; 673 AA.
 AC 09CGKO;
 DT 01-JUN-2001 (TREMBLrel. 17, Created)
 DT 01-JUN-2001 (TREMBLrel. 17, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Glycyl-tRNA synthetase beta chain (EC 6.1.1.14).
 GN GLYT OR LI1096.
 OS Lactococcus lactis (subsp. lactis) (Streptococcus lactis).
 OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
 OC Streptococcaceae; Lactococcus.
 OX NCBI_TaxID=1360;
 RN 11
 RP SEQUENCE FROM N.A.
 RC STRAIN=LI1403;
 RX MEDLINE=21235186; PubMed=11337471;
 RA Bojoltin A., Winkler P., Mauger S., Jallion O., Malarme K.,
 RA Weissenbach J., Ehrlich S.D., Sorokin A.;
 RT "The complete genome sequence of the lactic acid bacterium Lactococcus
 lactis ssp. lactis IL1403.";
 RT Genome Res. 11:731-753(2001).
 RL EMBL, AE006341; AAK05194.1; -
 DR InterPro: IPR002311; tRNA_synth_2f.
 DR Pfam: PF02092; tRNA_synth_2f; 1.
 DR PRINTS: PRO1045; TRNASYNTHGB.
 DR TIGRFAMS: TIGR00211; glys; 1.
 KW Aminocyl-tRNA synthetase; Complete proteome.
 SQ SEQUENCE 673 AA; 75717 MW; 99CF078804DAE69E CRC64;

Query Match 27.4%; Score 164; DB 16; Length 673;
 Best Local Similarity 38.0%; Pred. No. 2.2e-05;
 Matches 38; Conservative 21; Mismatches 37; Indels 4; Gaps 2;

OY 25 NLKKADALGEVNESLLOODEKALYAAAGLQPKIAAIVAEGNFRITLSELASVKKPOV 84
 Db 575 NLVKKK-KDIKEINPTLFEDAEALYNNVISLQNMWTYMPGEKFRATVHSLA---PAI 630
 OY 85 DAFPDGVMAEDAFAVQNRNLINRLAEQMNADIALL 124
 Db 631 ETFEESVMAEDLSVRDKRIALLSEVALTSVMADFSLL 670

Search completed: December 2, 2002, 12:02:13
 Job time : 12.8232 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 : Search time 36.1203 Seconds
(without alignments) 1283.801 Million cell updates/sec

Title: US-10-007-267-3
Perfect score: 1823
Sequence: 1 MQPLVSVLICAVNEKYFAQ.....RLIKNRQARSACKEGEI 348

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues
Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

A.Geneseq_101002.*
1: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1980.DAT:*
2: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1981.DAT:*
3: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1982.DAT:*
4: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1983.DAT:*
5: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1984.DAT:*
6: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1985.DAT:*
7: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1986.DAT:*
8: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1987.DAT:*
9: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1988.DAT:*
10: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1989.DAT:*
11: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1990.DAT:*
12: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1991.DAT:*
13: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1992.DAT:*
14: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1993.DAT:*
15: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1994.DAT:*
16: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1995.DAT:*
17: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1996.DAT:*
18: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1997.DAT:*
19: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1998.DAT:*
20: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA1999.DAT:*
21: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2000.DAT:*
22: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2001.DAT:*
23: /SIDS2/gcgdata/geneseq/geneseqp-emb1/AA2002.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1823	100.0	348	17	AA191311
2	1820	99.8	348	18	AAW06576
3	1709	93.7	346	23	AAU72923
4	1046	57.5	337	17	AA191314
5	1046	57.4	337	18	AAW06579
6	254.5	14.0	274	22	AA191315
7	244	13.4	274	22	AAW83330
8	227	12.5	278	21	AAV68976
9	226.5	12.4	334	21	AAV68963
10	225	12.3	298	22	AA191313

11	217.5	11.9	324	18	AAW14078
12	217.5	11.9	324	18	AAW22177
13	214.5	11.8	316	21	AAV54072
14	214.5	11.8	316	21	AAV43774
15	211	11.6	322	22	AA191314
16	210	11.5	322	21	AAV68974
17	210	11.5	965	23	AAV96213
18	210	11.5	965	23	AAW47335
19	210	11.5	965	23	AAW47336
20	208.5	11.4	972	20	AAV06212
21	208.5	11.4	972	21	AAV43099
22	207.5	11.4	336	21	AAV54098
23	207.5	11.4	336	21	AAV43800
24	205	11.2	329	22	AA191315
25	200.5	11.0	327	23	AA191316
26	200	11.0	702	21	AAV96212
27	195	10.7	301	21	AAV97206
28	194.5	10.7	322	21	AAV68975
29	193.5	10.6	268	21	AAV54093
30	193.5	10.6	268	21	AAV43795
31	191.5	10.5	674	19	AAW61238
32	191.5	10.5	674	23	ABP54656
33	190.5	10.4	302	21	AAV97213
34	190	10.4	270	21	AAV97203
35	189.5	10.4	350	22	AA191317
36	185.5	10.2	311	22	AB191318
37	183	10.0	332	21	AAV68962
38	181	9.9	328	21	AAV81720
39	179	9.8	345	22	AAV90140
40	179	9.8	389	21	AAV97202
41	177	9.7	322	21	AAV54073
42	177	9.7	322	21	AAV43773
43	175	9.6	270	19	AAW61237
44	175	9.6	270	23	ABP54655
45	175	9.6	281	19	AAW61236

ALIGNMENTS

RESULT 1	AA191311	standard; Protein; 348 AA.
ID	AA191311	
AC	AA191311	
DT	09-JUL-1996	(first entry)
DE	N. gonorrhoeae glycosyltransferase LgtA.	
KW	Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus; vaccine.	
OS	Neisseria gonorrhoeae strain F62.	
PN	MO9610086-A1.	
PD	04-APR-1996.	
PF	25-SEP-1995;	95WO-US12317.
PR	26-SEP-1994;	94US-0312387.
PA	(UNIQ) UNIV ROCKEFELLER.	
PI	Gotschlich EC;	
DR	WPI: 1996-200924/20.	
DR	N-PSDB; AAT14061.	
PT	Nucleic acids encoding glycosyl transferase(s) - used in the diagnosis of infection with Neisseria and for the biosynthesis of oligo:saccharide(s)	

xx Claim 8; Fig 2b; 81pp; English.

xx

xx 5 glycosyltransferases (AAR91311-15) are products of the 1gt locus

cc (AAR14061) of *Nesleria gonorrhoeae* strain F62. Glycosyltransferase

cc 1gtA (AAR91311) can be obt. by expression of the 1gtA coding

cc sequence in recombinant host cells. A method for adding GalNAc

cc or GlcNAc beta1-3 to Gal compounds contacting a reaction mixture

cc contg. activated GalNAc or GlcNAc to an acceptor moiety comprising

cc a Gal residue in the presence of 1gtA. Oligosaccharides can be

cc produced that, when attached to non-toxic lipids, are useful for

cc *Nesleria* vaccine prepn. Blood group core oligosaccharides, and

cc mimics of lacto-N-neotetraose, gangliosides and saccharide portions

cc of globoglycolipids can also be produced using the enzymes.

cc

Query Match	100.0%	Score 1823	DB 17	Length 348
Best Local Similarity	100.0%	Pred. No. 3.2e-184		
Matches 348; Conservative	0	Mismatches	0	Gaps 0

Qy	1	MOPVSLICAVNEXKXFPQSLAAVYNNQWRMLDILLYDQSGTGTGTAIKDKQKRSR	60
Qy	1	MOPVSLICAVNEXKXFPQSLAAVYNNQWRMLDILLYDQSGTGTGTAIKDKQKRSR	60
Db	1	MOPVSLICAVNEXKXFPQSLAAVYNNQWRMLDILLYDQSGTGTGTAIKDKQKRSR	60
Qy	61	KILAQNSGLIPSLNIGIDELAKSGGGGEXIAFTDADDIASPGWIEKIVGEMEKDR	120
Db	61	KILAQNSGLIPSLNIGIDELAKSGGGGEXIAFTDADDIASPGWIEKIVGEMEKDR	120
Qy	121	TAMGAMLEVIISSEKDGRRRLARHHKHKIKKPTREDDIAAFPPGPNTHNTYIMR	180
Db	121	TAMGAMLEVIISSEKDGRRRLARHHKHKIKKPTREDDIAAFPPGPNTHNTYIMR	180
Qy	181	DGGLRYDTERDMAEDYQFWYDYSKIGRLAYYPEALVKYRLAHANOVSSKSHV	240
Db	181	DGGLRYDTERDMAEDYQFWYDYSKIGRLAYYPEALVKYRLAHANOVSSKSHV	240
Qy	241	OKTANPDLQSMGKTPFDSLEYRQKKAAYELPEKDLPEEDFEARARELYQCKR	300
Db	241	OKTANPDLQSMGKTPFDSLEYRQKKAAYELPEKDLPEEDFEARARELYQCKR	300
Qy	301	PSGAMLDFAADGRMRRLFTLROYFGILYLRILKNRQARSDSAGKQEI	360
Db	301	PSGAMLDFAADGRMRRLFTLROYFGILYLRILKNRQARSDSAGKQEI	360

RESULT 2
AAW06576
ID AAW06576 standard; Protein; 348 AA.

DT 21-MAR-1997 (first entry)
XX
XX Neisseria polyglycosyltransferase.
DE
XX Neisseria polyglycosyltransferase: N-acetylglucosaminyl transferase.
KW Polyglycosyltransferase: N-acetylglucosaminyl transferase.
KW N-acetylglucosaminyl transferase; lipo-oligosaccharide.
XX
OS Neisseria gonorrhoeae ATCC 33084.

PN WO9640971-A1.

PD 19-DEC-1996.

AA 03-JUN-1996; 96WO-US08323.
PF

AA 07-JUN-1995; 95US-0478140.
PR

AA
PA
(NEOS-) NEOSE TECHNOLOGIES INC.

PI Buczala SL, Johnson KF, Roth S;

DR WPI; 1997-052351/05.
DR N-PSDB; AAT49230.
DR

PT Transfer of at least 2 saccharide units using
PT poly:glycosyl:transferase - isolated from *N. gonorrhoeae*, catalyses
PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
PT single galactose moiety

PS Disclosure; Fig 2A-C; 38pp; English.

A novel polylysyltransferase (PGTase) (AA06576) from *Neisseria gonorrhoeae* ATCC 3384 catalyses the stereospecific conjugation of 2 specific activated saccharide units (e.g., UDP-GlcNAc, UDP-GalNAc, UDP-Gal) to specific acceptors having a galactose moiety at a non-reducing end. It is the first PGTase reported to be capable of transfer of more than one different saccharide moiety. The PGTase is encoded by nucleotides 445-1468 of a lipopolysaccharide gene (AA049330). It can be produced in transformed host cells and used in oligosaccharide prodn.

Sequence 348 AA;

Query Match	99.8%	Score 1820;	DB 18;	Length 348;
Best Local Similarity	99.7%;	Pred. No. 6.6e-184;		
Matches 347; Conservative	1;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	MOPLVSVLICAYNVEKYTAQSLAAVYNQTNWRNDLILVDDGSTDGTATAKDEQOKRDSRI	60
	:		
Db	1	LOPLVSVLICAYNVEKYTAQSLAAVYNQTNWRNDLILVDDGSTDGTATAKDEQOKRDSRI	60
Qy	61	KILQAQNSGLIPSNITIGDELAKSSGGGGGYIARTADADLASGWIETKIVGEMKDRSI	120
	:		
Db	61	KILQAQNSGLIPSNITIGDELAKSSGGGGGYIARTADADLASGWIETKIVGEMKDRSI	120
Qy	121	IAMGALEVLSEKXGNRLRHHKHGKIMKRPETHEDIAEPFGNDIHNNTYIMRSTYI	180
	:		
Db	121	IAMGALEVLSEKXGNRLRHHKHGKIMKRPETHEDIAEPFGNDIHNNTYIMRSTYI	180
Qy	181	DGGLRVTDERDAEDYOEFMYDVSKIGRLAAYPEALVYRLHANOVSSKHSYRQHEIAGOI	240
	:		
Db	181	DGGLRVTDERDAEDYOEFMYDVSKIGRLAAYPEALVYRLHANOVSSKHSYRQHEIAGOI	240
Qy	241	OKTARNDFLOQSMGFRTREDSELEYROTKAAYELPEKDLPEEDFERARFYLXOCFKRTDP	300
	:		
Db	241	OKTARNDFLOQSMGFRTREDSELEYROTKAAYELPEKDLPEEDFERARFYLXOCFKRTDP	300
Qy	301	PGSAMUDFPAADGMRRLFTLROYFGILYRLIKNRQASDSAGDEORI	348
	:		
Db	301	PGSAMUDFPAADGMRRLFTLROYFGILYRLIKNRQASDSAGDEORI	348

RESULT 3
AAU72923
ID AAU72923 standard; Protein; 346 AA.

AA AU72923; AC

DT 12-MAR-2002 (first entry)

aa Neisseria meningitidis virulence protein #13

AA Meningitis; virulence; gene; antibacterial; vaccine; veterinary;
KW

	XX
XXXXXXXXXXXXXXX	MY
XXXXXXXXXXXXXXX	NY
XXXXXXXXXXXXXXX	TX
XXXXXXXXXXXXXXX	VA
XXXXXXXXXXXXXXX	WI
XXXXXXXXXXXXXXX	WY

US NEISSEL ID MENTING LILARS
XX

PN WO2001053112
XX

PD 13-NOV-2001.
XX

XX
PE
06-MAY-2001, 2001MC 08020003

प्र

Query Match	Best local similarity	94.0%	Pred. No. 3.8e-172;	Matches 327;	Conservative 11;	Mismatches 8;	Indels 2;	Gaps 1;
QY 1	MQPLVSVLICAYNVEKYEAQSLAAVYVNOTWRMLDILIVDDGSTDGLTALAKDFOKRDSRI 60	1	LOPLVSVLICAYNVEKYEAQSLAAVYVNOTWRMLDILIVDDGSTDGLTALAKDFOKRDSRI 60					
QY 61	KTLAQAQNSGLIPSLNIGDELAKSGGGEGEYIARTDADDLASPGMIKIVGEMEKDRSI 120	61	KTLAQAQNSGLIPSLNIGDELAKSGGGEGEYIARTDADDLASPGMIKIVGEMEKDRSI 120					
QY 61	KTLAQAQNSGLIPSLNIGDELAKSGGGEGEYIARTDADDLASPGMIKIVGEMEKDRSI 118	61	KTLAQAQNSGLIPSLNIGDELAKSGGGEGEYIARTDADDLASPGMIKIVGEMEKDRSI 118					
QY 121	IAMGAWLEVLSEKQGNRLARHHKKGKTIKKTRREDIYAAPFGPNINNTMIMRSTY 180	121	IAMGAWLEVLSEKQGNRLARHHKKGKTIKKTRREDIYAAPFGPNINNTMIMRSTY 180					
QY 119	IAMGAWLEVLSEKQGNRLARHHKKGKTIKKTRREDIYAAPFGPNINNTMIMRSTY 178	119	IAMGAWLEVLSEKQGNRLARHHKKGKTIKKTRREDIYAAPFGPNINNTMIMRSTY 178					
QY 181	DGGLRYDTERDWAEDYQEFWYDVSKLGRLAYYPEALVKYFLHANQVSSKSHVROHEIAGCI 240	181	DGGLRYDTERDWAEDYQEFWYDVSKLGRLAYYPEALVKYFLHANQVSSKSHVROHEIAGCI 240					
QY 179	DGGLRYDTERDWAEDYQEFWYDVSKLGRLAYYPEALVKYFLHANQVSSKSHVROHEIAGCI 238	179	DGGLRYDTERDWAEDYQEFWYDVSKLGRLAYYPEALVKYFLHANQVSSKSHVROHEIAGCI 238					
QY 241	OKTARNDELQSGKGYTRDPSLEYKOTKAAAYLPEKDEEEDFERARRFLYOCFKRTDP 300	241	OKTARNDELQSGKGYTRDPSLEYKOTKAAAYLPEKDEEEDFERARRFLYOCFKRTDP 300					
QY 239	OKTARNDELQSGKGYTRDPSLEYKOTKAAAYLPEKDEEEDFERARRFLYOCFKRTDP 298	239	OKTARNDELQSGKGYTRDPSLEYKOTKAAAYLPEKDEEEDFERARRFLYOCFKRTDP 298					
QY 301	PGAWLDFPADGMRRLFTLROYFGILVRLIKNRRQARSDSAGKDEI 348	301	PGAWLDFPADGMRRLFTLROYFGILVRLIKNRRQARSDSAGKDEI 348					
QY 299	PGAWLDFPADGMRRLFTLROYFGILVRLIKNRRQARSDSAGKDEI 346	299	PGAWLDFPADGMRRLFTLROYFGILVRLIKNRRQARSDSAGKDEI 346					
RESULT 4								
AA91314	AA91314 standard; Protein: 337 AA.							
AA91314	AA91314;							
09-JUL-1996	(first entry)							
N. gonorrhoeae	glycosyltransferase Igtd.							
glycosyltransferase; lipo-	oligosaccharide; lgt gene; LOS locus;							
vaccine.								
Neisseria gonorrhoeae	strain F62.							
WO9610086-A1								

XX	04-APR-1996.
PD	
XX	
PF	25-SEP-1995; 95WO-US12317.
XX	
PR	26-SEP-1994; 94US-0312387.
XX	
PA	(UYRQ) UNIV ROCKEFELLER.
XX	
PI	Gotschlich EC;
XX	
DR	WPI; 1996-200924/20.
XX	
DR	N-PSDB; AAT14061.
XX	
PT	Nucleic acids encoding glycosyl transferase(s) - used in the
PT	diagnosis of infection with Neisseria and for the biosynthesis of
PT	oligo:saccharide(s)
XX	
PS	Claim 11; Fig 2e; 81pp; English.
XX	
XX	5 glycosyltransferases (AAR91311-15) are products of the lgt locus
CC	(AAT14061) of Neisseria gonorrhoeae strain F62. Glycosyltransferase
CC	LgtD (AAR91314) can be obt'd. by expression of the lgtD coding
CC	sequence in recombinant host cells. A method for adding GalNAc
CC	or GLcNAc beta1-3 to Gal comprises contacting a reaction mixture
CC	contg. activated GalNAc or GLcNAc to an acceptor moiety comprising a
CC	Gal residue in the presence of LgtD. Oligosaccharides can be produced
CC	that, when attached to non-toxic lipids, are useful for Neisseria
CC	vaccine prep'n. Blood group core oligosaccharides, and mimics of
CC	lacto-N-neotetraose, gangliosides and saccharide portions of
CC	globoglycolipids can also be produced using the enzymes.
XX	
Sequence	337 AA:

	Query Match	57.5%	Score 1049:	DB 17:	Length 337:	
	Best Local Similarity	64.5%	Pred. No. 2.9e-102:			
	Matches 214;	Conservative 33;	Mismatches 83;	Indels 2;	Gaps 1	
OY	1 MQLPVSVLICAAVNRKVFQASLAAVVNQWNRNLDILIVDGSSTDGTLAIYAKDFOKRDSRI 60					
Db	1 MQPLVSVLICAVNAEKYFQAQSLAAYVGQWNRNLDILIVDGSSTDGTPALAHHQEDGRL 60					
OY	61 KLLAQONSGLLPSLNIGLDELAKSGGGGGGYIARTDADDDIASPGNIETKIVGEMEKDRSI 120					
Db	61 RLISPNRNIGFLASINIGDELAQS--GGGEYIARTDADDIASPGNIETKIVGEMEKDRSI 118					
OY	121 IAMGMLEVLSEKQGNRLRAHHKKTIWKPTREHDIAPFPFGNPINNTIMRRSYI 180					
Db	119 IAMGMLEVLSEENKNSVIATAARNALMDKPTRHEIDIAYVPFGNPINNTIMRRSYI 178					
OY	181 DGLGRDYTERDAEDYQEFNYDVSKLGRLLYVEPALVKYLIAHNOVSSKSASHVRQETHAOGI 240					
Db	179 DGGLRFDPAYIHAEYIKFFYEAGKIGRLRAYTEALVKYKFHDQGISKYNLQORRTAMKI 238					
OY	241 OKTARNDLIQSKGETREDLSLEYROTKAAAYELPEKDLPEDPERARRFLKYOCFKFTDP 300					
Db	239 KEIRIAGYKKAAGIIVGAQCNLNGLLKSAVALYERALKSGODIGCLRFLYEYFLSLEYK 298					
OY	301 PEGANUDFRADGRMRRLFTLRYPFGILYLIK 332					
Db	299 SLTDLDLFLLTRVMRKRLFAAPQRILTKRMRL 330					
	RESULT 5					
	AAM06579					
XX	AAM06579 standard; Protein; 337 AA.					
XX	AC					
XX	AA006579;					
XX	21-MAR-1997 (first entry)					
XX	Lipo-oligosaccharide gene-encoded protein.					
DE						
XX						

KW Polyglycosyltransferase: N-acetylglucosaminyl transferase;
KW N-acetylglucosaminyl transferase; lipo-oligosaccharide.
XX
OS Neisseria gonorrhoeae ATCC 33084.
XX
PN W09640971-A1.
XX
PD 19-DEC-1996.
XX
PE 03-JUN-1996; 96MO-US08323.
XX
PR 07-JUN-1995; 95US-0478140.
XX
PA (NEOS-) NEOSE TECHNOLOGIES INC.
XX
PI Bucala SL, Johnson KF, Roth S;
XX
DR WPI: 1997-052351/05.
DR N-PSDB; AAT49230.
XX
PT Transfer of at least 2 saccharide units using
PT poly:glycosyl:transferase - isolated from N. gonorrhoeae, catalyses
PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
PT single galactose moiety
XX
PS Disclosure; Fig 2E-F; 38pp; English.
XX
CC A lipooligosaccharide-encoding gene region (AAT49230) of Neisseria
CC gonorrhoeae ATCC 33084 includes coding sequences for 5 proteins
CC (AAW06576-80), one of which (AAW06576) is a polyglycosyltransferase
CC that catalyses the addition of GlcNAc and GalNAc disaccharides to
CC a galactose moiety. The function of the other 4 proteins is not
CC stated in the specification.
XX
SQ Sequence 337 AA;

Query Match 57.4%; Score 1046; DB 18; Length 337;
Best Local Similarity 64.2%; Pred. No. 6.1e-102;
Matches 213; Conservative 34; Mismatches 83; Indels 2; Gaps 1;

OY 1 MPTLVSVLICAVNVEKFFQSLAAVNOVNRNDLILYVDSSTGTLTAIKPOKRSRI 60
DB 1 LQPLVSVLICAVNAEYFAQSLAAVVGOTWRNDLILYVDSSTGTLTAIRHFOEDGRI 60
OY 61 KIILAOQNSGLIPSLINIGLDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKRSI 120
DB 61 RIISNPRNIGFTASLNTIGDELAKS--GGGEYIARTDADDIASPGWIEKIVGEMEKRSI 118
OY 121 IAWGAWLEVLSEKDGNRRLARRHHKGIWKKPTRHEDIAAFPEFGNPINNTMIRRSYI 180
DB 119 IAWGAWLEVLSEKDGNRRLARRHHKGIWKKPTRHEDIAAFPEFGNPINNTMIRRSYI 178
OY 181 DGLRPTDERMAEDQOFYVDVSKIGRLAYVPEALVKYRFLANOVSSKSHVROHETAGSI 240
DB 179 DGLRPTDPAYIHAEDEKFEYKGLRGLAYVPEALVKYRHOQVSSKTNLOQRRAWKI 238
OY 241 OKTARDFLOSMGFKTRFSLSEYRQTAAAYELPEKDLPEDEFERARALYOCFKRTDPP 300
DB 239 KEIRIRGYNKAGINAGACLOINGLKSTAYALYKALSGODIGCLRLFLYEYFLSEKY 298
OY 301 PSGAWLDPADGMRRLFTLRQYFGILYRLIK 332
DB 299 SLTDLDLFLTRVYMRKLFAPAOYRKILKMLR 330

RESULT 6
AAG90151
ID AAG90151 standard; Protein; 274 AA.
XX
AC AAG90151;
XX
DT 26-SEP-2001 (first entry)
XX

DE C glutamicum protein fragment SEQ ID NO: 3905.

XX Coryneform bacterium; amino acid synthesis; vitamin; saccharide;
KW organic acid synthesis.
XX
OS Corynebacterium glutamicum.
XX
PN EP1108790-A2.
XX
PD 20-JUN-2001.
XX
PE 18-DEC-2000; 2000EP-0127688.
XX
PR 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX

(KYOW) KYOMA HAKKO KOGYO KK.

XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX

DR WPI: 2001-376931/40.
DR N-PSDB; AAH65370.

PT Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
XX
PS Claim 17; SEQ ID NO: 3905; 246pp + Sequence listing; English.

CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.

XX
SQ Sequence 274 AA;

Query Match 14.0%; Score 254.5; DB 22; Length 274;
Best Local Similarity 30.0%; Pred. No. 3e-18;
Matches 80; Conservative 49; Mismatches 91; Indels 47; Gaps 10;

OY 3 PLVSVLICAVNVEKFFQSLAAVNOVNRNDLILYVDSSTGTLTAIKD--FOKRSR 59
DB 6 PLVSVIPIAIDEKCSGSIKSCVENYENMOIVLVDS-----APIKDVQWKEHR 59
OY 60 KIILAOQNSGLIPSLINIGLDELANSGGGGGEYIARTDADDIASPGWIEKIVGEMEKRS 119
DB 60 IRIVEQRIKRGPTSLNIGIK-----ASDGLIARLSDDDLAASRLSKOEFELRNHY 113
OY 120 IAWGAWLEVLSEKDGNRRLARRHHKGIWKK-----PTRHEDIAAFPEFGNPINNTMIR 175
DB 114 IICVATKTHINE-----HGKIFGQSADLPT--SDIRQTLVKNPILHSSVNY 160
OY 176 RRSVID--GGLRPTDERMAEDQOFYVDVSKIGRLAYVPEALVKYRFLANOVSSKSHS-- 230
DB 161 RKQVVRQIGG--YSLMTSPQSDYELFLRLSKIGALIGLDESLSSYRIHGGQSRKTSPPK 218
OY 231 -----VRQHEIAGIOKT--ARNDFL 249
DB 219 KTWIILKRRMELASFLKRSVPVQIFL 245

xx	AAW89330	standard; peptide: 50 AA.
xx	AAW89330;	
xx	26-FEB-1999	(first entry)
xx	Neisseria meningitidis IgtA C-terminal peptide.	
xx	Neisseria meningitidis IgtC; IgtB; beta-1,4-galactosyltransferase;	
xx	glycosyltransferase; proteolytic enzyme.	
xx	Neisseria meningitidis.	
xx	MO9854331-A2.	
xx	03-DEC-1998.	
xx	26-MAY-1998;	98WO-IB00975.
xx	27-MAY-1997;	97US-0047751.
xx	(CANADA) NAT RES COUNCIL CANADA.	
xx	Wakarchuk WW, Young NM;	
xx	WPI; 1999-035177/03.	
xx	Expressing high levels of glycosyltransferases - comprises use of	
xx	either host cells deficient in proteolytic enzymes or modified	
xx	glycosyltransferase genes deleted in a proteolytic recognition site	
xx	Example 1; Fig 8; 61np; English.	
xx	A method has been developed of expressing a glycosyltransferase in a	
xx	host cell. The method comprises introducing into the host cell a nucleic	
xx	acid encoding the glycosyltransferase and incubating the host cell under	
xx	conditions appropriate for expression of the glycosyltransferase, where	
xx	the host cell substantially lacks a protease that cleaves polypeptides	
xx	between two consecutive positively charged amino acid residues. The	
xx	glycosyltransferase can be used in vitro production of	
xx	oligosaccharide structures which are potential therapeutic agents for	
xx	use in the manipulation of cell-cell recognition events, particularly	
xx	adhesion of bacteria and viruses to mammalian cells and leukocyte-	
xx	endothelial cell interaction through selectins in inflammation. The	
xx	method provides more readily recoverable active glycosyltransferases	
xx	than prior art methods involving mammalian glycosyltransferases. The	
xx	present sequence represents a C-terminal peptide from Neisseria	
xx	meningitidis IgtA from the present invention.	
xx	Sequence 50 AA:	
xx	Query Match 13.4%; Score 244; DB 20; Length 50;	
xx	Best Local Similarity 91.8%; Pred. No. 3.1e-18;	
xx	Matches 45; Conservative 3; Mismatches 1; Indels 0; Gaps 0;	
Oy	286 ARRFYOCFKRTPPGSGAWLDFADGRRRRTFTROYRGILYRLIKNR 334	
Oy		
Db	2 ARRFYOCFKRITDLPAGAWLDFADGRRRRLFTLRKYFGILLRLIKNR 50	
xx	RESULT 8	
xx	AAW8976	
xx	AAW8976 standard; Protein: 278 AA.	
xx	AAW8976;	
xx	30-MAY-2000 (first entry)	
xx	Cpsik protein which has glycosyltransferase activity.	
xx	Capsular gene cluster; serotype 1; polysaccharide biosynthesis;	

XX	capsular component; antigen; regulation; chain length determination;
KW	complement-mediated opsonophagocytosis; serotype-specific detection;
RN	antigen; vaccine; Streptococcal disease; CpsII; CpsI; CpsIG.
KV	CpsIH; CpsII; CpsID; CpsIK; glycosyltransferase; CP polymerase.
XX	
OS	Streptococcus suis.
XX	
PN	WO200005378-A2.
XX	
PD	03-FEB-2000.
XX	
PE	19-JUL-1999; 99WO-NL00460.
XX	
PR	22-JUL-1998; 98EP-0202465.
XX	
PA	(DIEN-) STRICHTING DIENST LANDBOUWKUNDIG ONDERZOE.
XX	
P1	Smith HE;
DR	WPI: 2000-195104/17.
XX	
N	PSDB; AAZ60930.
XX	
PT	New nucleic acid containing the capsular gene cluster of Streptococcus
PT	suis, used for serotype-specific detection and to generate antigens or
XX	mutants for vaccination
XX	
PS	Disclosure; Fig 4; 14pp; English.
XX	
CC	The proteins AAY68970-76 are encoded by the capsular gene cluster of
CC	Streptococcus suis serotype 1. The genes in this cluster are involved
CC	in polysaccharide biosynthesis of capsular components and antigens. The
CC	proteins have glycosyltransferase activities (CpsIF, CpsIG, CpsII,
CC	CpsIJ, CpsIK) and CP polymerase activities (CpsIH). The capsule confers
CC	bacterium resistance to complement-mediated opsonophagocytosis. The
CC	gene cluster is used as a source of probes and primers for
CC	serotype-specific detection of S. suis and is also useful for
CC	recombinant production of the proteins. The proteins are then useful
CC	for producing antigens that can be used in vaccines, for controlling
CC	or eradicating a Streptococcal disease, in humans or animals,
XX	e.g. against S. suis in pigs.
XX	
SQ	Sequence 278 AA;
Query Match	12.5%; Score 227; DB 21; Length 278;
Best Local Similarity	24.6%; Pred No. 2.5e-15;
Matches	68; Conservative 60; Mismatches 88; Indels 60; Gaps
OY	5 VSLICAVNVEKYFAQSILAAVVNQTNRNDILIVDGSSTDGTALAKDFOKRDSRIKILA 64
Db	: : : : : : : : : : : : : : : : : : : :
	7 ISIIPIYNVEKKLCIDSIVNQTYTKHEILLVDNGSDNSEECILAAVKDDIRRYF- 65
OY	65 QAONGSLPNTLGDELAKSGGGEGEYIARTDADDIASPGMIEKTVGEMEKDRSIIANG 124
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : :
	66 KKEENGSLSPARNYGISR-AK-----GDYLAFIDSDPFHSEFIQRLEAEIERENALVAVA 119
OY	125 AWLEYLSEKEDGNRLARHHKHGKIWMKPPRHEDIAPFFPGAPIHNNTIMMR-RSYIDGG 183
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : :
	120 CYDRV-----DASGHFLTAERPLPTNOAVALSRNNCKKL 152
OY	184 LRYDTER-----DMAEDYOF-----NYDVSKGLGRLAYYPEALVKYRL 220
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : :
	153 LEADGHRFVACNKLKYKLFDPDFREKCIHEDEYFRLLYLELKAIVNECLUYVVD 212
OY	221 HANOVSSKHSVRQHEIAOCIQ-KTARNDFLOSMEGR 255
Db	: : : : : : : : : : : : : : : : : : : : : : : : : : : :
	213 RENSITTS-SMTDHRFHCLLEFONERMDFEYSRGDK 247
RESULT 9	
AAY68963	
ID	AAY68963 standard; Protein; 334 AA.
XX	

[illegible]

Db	210	RENSIIITSSMTDHRHCLE--FQNERMFYESRG-----DKEL	246
Oy	279	PEDEFERARAFYOCFKRPTDPPSGAWLDFAA-----DGRMRLETLRQYFGILY	328
Db	247	-----LLECYR-----SFLAFVAVLFLGKYNHMLSKQCKL--LQTLFRIVY	285
Oy	329	RLIK-NRROA 337	
Db	286	KOLKONKRLA 295	
RESULT 10			
AAB96313			
ID	AAB96313	standard; Protein; 298 AA.	
XX			
AC	AAB96313;		
XX			
DT	29-OCT-2001	(first entry)	
XX			
DE	putative glycosyltransferase, involved in cell wall biogenesis #1.		
XX			
KW	Hyperthermophilic archaeon; hyperthermophilic protein.		
XX			
OS	Pyrococcus abyssi.		
XX			
PN	FR2792651-A1.		
PD	27-OCT-2000.		
XX			
XX	21-APR-1999;	99FR-0005034.	
XX			
PR	21-APR-1999;	99FR-0005034.	
XX			
PA	(CNRS) CNRS CENT NAT RECH SCT.		
PA	(IFREMER) IFREMER INST FR RECH EXPL MER.		
XX			
PI	Forterre P, Thierry JC, Prieur D, Dietrich J, Lecompte O;		
PI	Querellou J, Weissenbach J, Saurin W, Hellig R;		
XX			
DR	WPI, 2001-126236/14.		
XX			
XX	New nucleotide sequences isolated from Pyrococcus abyssi encode		
PT	proteins useful in industry -		
XX			
PS	Claim 7; Pages 981-982; 1657pp; French.		
XX			
CC	The present invention relates to the genomic sequence of Pyrococcus		
CC	abyssi (see AAF84431 and AAH41223-7) and P. abyssi proteins. P. abyssi is		
CC	a hyperthermophilic archaeon, which is isolated from deep-sea		
CC	hydrothermal vents. The present sequence is one such P. abyssi protein.		
CC	The proteins of the present invention have various potential industrial		
CC	uses, since the proteins are stable at very high temperatures, some up to		
CC	110 degrees centigrade.		
CC	Note: This patent is in the same patent family as WO200065062, which		
CC	contains additional sequences as shown in AAB99132-AAB99143,		
CC	AAH75903-AAH75920 and AAG66436.		
CC			
XX			
XX	Sequence	298 AA;	
Query Match	12.3%;	Score 225;	DB 22;
Best Local Similarity	26.5%;	Pred. NO. 4,6e-15;	
Matches	65;	Conservative	52;
		Mismatches	96;
		Indels	32;
		Gaps	8
Oy	2	OPVLVLCATVVEKTFPAOSLAAYVNOQWRNIDILVDDGSDGTGLAIKKDFOKRDSRIK	61
Db	3	REIVSVIIPTYNRAMLLRRAIVSYLNQKFKDELIVDDASDNPPEVYESIE--DGRIR	60
Oy	62	ILAAQONSLIPSLINIGDELAKSGGCGEYIARPDADDIAPGMIETKIVGEME---KRR	118
Db	61	YIRLRKNSGGPIARNIGIKK-AK-----GRFALLDDDEWLPPIRLVGVVRKFENIGKEF	114
Oy	119	SIIAAGVLEVISEKKDGNRLARHHRHGKIVTKPTRHEDIAAFPPFGNPIHNNTMTIMRS	178

CC eps8, eps9 and eps10) that are involved in the biosynthesis of
CC exopolysaccharides (EPS). The enzymes catalyse the formation of
CC specific intersugar bonds. The enzymes catalyse a process that includes
CC at least one step of forming a bond (in alpha or beta anomeric form)
CC between C1, carrying the reducing aldehyde group of an activated D-Galp
CC (galactose in pyranose form), and a phosphate group on a lipophilic or
CC antilumior or probiotic properties or are used in foodstuffs to improve
CC organoleptic qualities and flavour. When expressed by lactic acid
CC bacteria, EPS impart a free-flowing character and/or a smooth, creamy
CC texture to acidified milk products (yoghurt or cheese).

SQ Sequence 316 AA;

Query Match 11.8%; Score 214.5; DB 21; Length 316;
Best Local Similarity 35.7%; Pred. No. 6.5e-14;
Matches 45; Conservative 29; Mismatches 45; Indels 7; Gaps 2;

OY 1 MOPVSLICAVNVEKFAOSLAAYVNOTWRNLDILIVDGSSTGTLAIKDFOKRDSRI 60
DB 1 MNPILSIIVPYNVEKIRTCIESILAQTYRNIEVILVNDGSTDSLAVISDLICSHNI 60
OY 61 KTLAQONSLIPSLNIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
DB 61 KVINQ-KNOGLSVARNLTGID-----AATGKYIAFYADDDKIKPFVSSLYOTADKTGAD 113
OY 121 IAMGAM 126
DB 114 IYRGSF 119

RESULT 15
AAB47426
ID AAB47426 standard; Protein; 332 AA.

XX AAB47426;

XX 17-OCT-2001 (first entry)

XX EpsN.

KW EPS plasmid; Lactococcus lactis subspecies cremoris Ropy352;
KM exopolysaccharide; EPS352; milk; thickener; glycosyltransferase;
KW biosynthesis; fruit juice.

OS Lactococcus lactis.

PN WO200157234-A2.

PD 09-AUG-2001.

PF 02-FEB-2001; 2001WO-US03404.

PR 02-FEB-2000; 2000US-0179888.

PR 16-OCT-2000; 2000US-0241098.

PA (UYOR-) UNIV OREGON STATE.

PA (USDA) US DEPT OF AGRICULTURE.

PT Trempey JE, Knoshaug EP, Sandline WE, Ahlgren JA, Dierksen KP;

DR WPI; 2001-488889/53.

DR N-PSDB; AAH43198.

PT New bacterium useful in pharmaceutical formulations, food products and
PT beauty cosmetics, comprises characteristics of Lactococcus lactis
PT cremoris Ropy 352

PS Claim 28; Page 67-68; 73pp; English.

CC The sequences given in AAB47426-27 are encoded by a fragment of the EPS
CC plasmid derived from L. lactis subspecies cremoris Ropy352. The EPS
CC plasmid is about 32 kb in size and encodes at least 13 active genes.

CC The enzymes encoded by these genes allow the bacteria to produce an
CC exopolysaccharide, designated EPS352. When EPS352 is expressed in or
CC added to milk, it imparts highly desirable sensory characteristics to
CC the milk, including making the milk very thick, with a very smooth
CC mouth-feel, and slightly sweet with an obvious chewable-bite. Open
CC reading frames (ORF's) M and N show homology to glycosyltransferase
CC involved in EPS352 biosynthesis. L. lactis cremoris Ropy352 is
CC deposited with the USDA-ARS-NCAUR-NRRL as deposit accession number
CC NRRL B-30229. EPS352 is useful for thickening a liquid selected from
CC milk, a milk-based liquid, a whey-based liquid, a soy-based liquid,
CC and a fruit-juice. It is also useful as an additive in pharmaceutical
CC products, beauty care products and coating agents.

SQ Sequence 332 AA;

Query Match 11.6%; Score 211; DB 22; Length 332;
Best Local Similarity 24.5%; Pred. No. 1.6e-13;
Matches 60; Conservative 51; Mismatches 84; Indels 50; Gaps 7;

OY 4 LVSVLICAVNVEKFAOSLAAYVNOTWRNLDILIVDGSSTGTLAIKDFOKRDSRIKL 63
DB 5 LSIIVPYNSEKILRAAHSILNLOTYONIEVILVNDGSTDSQELISSFOKKDKRIK-L 63
OY 64 AQONSLIPSLNIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIAM 123
DB 64 YMKNLGVSHARNYGLDR-----ASGYIMEFLDPDDTYDKSYCLEMIGLKNKNADVYM 117
OY 124 GAM-----LE---VISEKD-GNRLARHKIKTKKTRIEDIAAFPP 163
DB 118 SNVYICKGNIPYNNNDLLECEGLSRDRTMSILSDTGFGEFW----- 163
OY 164 FGNPIHNTMIMRSYIDGLRYDTERDAEDYOFYVDVSKGLRAYYPEALVKRYLHAN 223
DB 164 -----TRIRKKNVIN-NVKFNESINVLEDMLFNISIVNHARIATYINKKHYPLORE 214
OY 224 QVSSK 228
DB 215 DSASK 219

Search completed: December 2, 2002, 11:58:59

Job time : 38.1203 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 12.9535 Seconds
(without alignments)
790.458 Million cell updates/sec

Title: US-10-007-267-3

Perfect score: 1823

Sequence: 1 MGPLSVLLICAVNVEKYFAQ.....RLIKNRQARSDSAGEDEI 348

Scoring table:

BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 10%

Listing first 45 summaries

Database : Issued_Patents_AA:*

- 1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
- 2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
- 3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
- 4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
- 5: /cgn2_6/ptodata/1/1aa/PCUTOS.COMB.pep:*
- 6: /cgn2_6/ptodata/1/1aa/Backfile1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1823	100.0	348	1	US-08-312-387B-3
2	1823	100.0	348	1	US-08-683-426-3
3	1823	100.0	348	1	US-08-683-458-3
4	1823	100.0	348	2	US-08-878-360-3
5	1823	100.0	348	4	US-09-333-412-3
6	1820	99.8	348	1	US-08-312-387B-11
7	1820	99.8	348	1	US-08-683-426-11
8	1820	99.8	348	1	US-08-683-458-11
9	1820	99.8	348	2	US-08-878-360-11
10	1820	99.8	348	3	US-08-478-140B-3
11	1820	99.8	348	4	US-08-478-140B-8
12	1820	99.8	348	4	US-09-333-412-11
13	1820	99.8	348	4	US-09-338-943-3
14	1820	99.8	348	4	US-09-338-943-8
15	1820	99.8	348	1	US-08-312-387B-5
16	1820	99.8	348	1	US-08-683-426-5
17	1820	99.8	348	1	US-08-683-458-5
18	1820	99.8	348	2	US-08-878-360-5
19	1820	99.8	348	2	US-09-333-412-5
20	1820	99.8	348	1	US-08-312-387B-12
21	1820	99.8	348	1	US-08-683-458-12
22	1820	99.8	348	1	US-08-878-360-12
23	1820	99.8	348	2	US-08-478-140B-5
24	1820	99.8	348	2	US-09-333-412-12
25	1820	99.8	348	4	US-09-338-943-5
26	1820	99.8	348	4	US-08-597-236-10
27	1820	99.8	348	1	US-08-597-236-10

28	217.5	11.9	324	1	US-08-746-682A-10	Sequence 10, Appl
29	210	11.5	965	4	US-09-437-277-3	Sequence 3, Appl
30	200	11.0	702	4	US-09-437-277-1	Sequence 1, Appl
31	191.5	10.5	674	4	US-08-961-083-200	Sequence 200, App
32	175	9.6	270	4	US-08-961-083-198	Sequence 198, App
33	175	9.6	281	4	US-08-961-083-196	Sequence 196, App
34	154.5	8.5	79	4	US-08-858-207A-521	Sequence 521, App
35	147.5	8.1	93	4	US-08-961-083-168	Sequence 168, App
36	130	7.1	358	4	US-09-134-001C-5633	Sequence 5633, Ap
37	129	7.1	303	4	US-08-961-083-202	Sequence 202, App
38	124.5	6.8	727	4	US-09-134-001C-4067	Sequence 4067, Ap
39	113.5	6.2	517	2	US-08-967-508-19	Sequence 19, Appl
40	113.5	6.2	517	2	US-08-967-506-19	Sequence 19, Appl
41	113.5	6.2	517	5	PCR-US94-02552-19	Sequence 19, Appl
42	113.5	6.2	559	2	US-08-967-508-9	Sequence 9, Appl
43	113.5	6.2	559	2	US-08-967-506-9	Sequence 9, Appl
44	113.5	6.2	559	5	PCR-US94-02552-9	Sequence 9, Appl
45	112.5	6.2	418	4	US-09-134-001C-4051	Sequence 4051, Ap

ALIGNMENTS

```
RESULT 1
US-08-312-387B-3
; Sequence 3, Application US/08312387B
; Patent No. 5545553
;
GENERAL INFORMATION:
;
APPLICANT: Gotschlich, Emil C.
;
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
;
NUMBER OF SEQUENCES: 12
;
CORRESPONDENCE ADDRESS:
;
ADDRESSEE: Klauber & Jackson
;
STREET: 411 Hackensack Avenue
;
CITY: Hackensack
;
STATE: New Jersey
;
COUNTRY: USA
;
ZIP: 07601
;
COMPUTER READABLE FORM:
;
MEDIUM TYPE: Floppy disk
;
COMPUTER: IBM PC compatible
;
OPERATING SYSTEM: PC-DOS/MS-DOS
;
SOFTWARE: PatentIn Release #1.0, Version #1.25
;
CURRENT APPLICATION DATA:
;
APPLICATION NUMBER: US/08/312, 387B
;
FILING DATE: July 7, 1994
;
CLASSIFICATION: 435
;
ATTORNEY/AGENT INFORMATION:
;
NAME: Jackson Esq., David A.
;
REGISTRATION NUMBER: 26,742
;
REFERENCE/DOCKET NUMBER: 600-1-095
;
TELECOMMUNICATION INFORMATION:
;
TELEPHONE: 201 487-5800
;
TELEFAX: 201 343-1684
;
TELEX: 133521
;
INFORMATION FOR SEQ ID NO. 3:
;
SEQUENCE CHARACTERISTICS:
;
LENGTH: 348 amino acids
;
TYPE: amino acid
;
TOPOLOGY: linear
;
MOLECULE TYPE: protein
;
US-08-312-387B-3
;
Query Match 100.0%; Score 1823; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 5.5e-150;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 MGPLSVLLICAVNVEKYFAQSLAAVYNQWRNLDLIYVDSGTDGTLATAKPKQRDSRI 60
|||||
DB 1 MGPLSVLLICAVNVEKYFAQSLAAVYNQWRNLDLIYVDSGTDGTLATAKPKQRDSRI 60
|||||
QY 61 KIIAQNNGSLIPISINIGHDELAKSGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120

```
|||||
Db 61 KIIA0A0NSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGMIETKIVGEMEKDSRI 120
QY 121 IAMGAWLEVLSEKDGKGNRLARHHKGIWKKPTRHEDIAAFPPFGNPJHNTMIMRSYI 180
Db 121 IAMGAWLEVLSEKDGKGNRLARHHKGIWKKPTRHEDIAAFPPFGNPJHNTMIMRSYI 180
QY 181 DGLRLDTERDMAEDYQFYWDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEITAGI 240
Db 181 DGLRLDTERDMAEDYQFYWDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEITAGI 240
QY 241 OKTARNDFLOSGMFKTRFDSLEYROTAKAAYELPEKDLPEEDFERARRELYOCFKRTDP 300
Db 241 OKTARNDFLOSGMFKTRFDSLEYROTAKAAYELPEKDLPEEDFERARRELYOCFKRTDP 300
QY 301 PSGAWLDFADGRMRRLFTLROYFGILYRLIKNRROARSDSAGKEDEI 348
Db 301 PSGAWLDFADGRMRRLFTLROYFGILYRLIKNRROARSDSAGKEDEI 348
```

RESULT 2

```
US-08-683-426-3
; Sequence 3, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683.426
; FILING DATE:
; CLASSIFICATION: 536
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-426-3
```

```
Query Match 100.0%; Score 1823; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 5.5e-190;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MOPIVSVLICAVNVEKYFAQSILAAVYVNTWRNLDILIVDGSIDGTLAIAXDFQKRDSTRI 60
Db 1 MOPIVSVLICAVNVEKYFAQSILAAVYVNTWRNLDILIVDGSIDGTLAIAXDFQKRDSTRI 60
```

```
QY 61 KIIA0A0NSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGMIETKIVGEMEKDSRI 120
Db 61 KIIA0A0NSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGMIETKIVGEMEKDSRI 120
QY 121 IAMGAWLEVLSEKDGKGNRLARHHKGIWKKPTRHEDIAAFPPFGNPJHNTMIMRSYI 180
Db 121 IAMGAWLEVLSEKDGKGNRLARHHKGIWKKPTRHEDIAAFPPFGNPJHNTMIMRSYI 180
QY 181 DGLRLDTERDMAEDYQFYWDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEITAGI 240
Db 181 DGLRLDTERDMAEDYQFYWDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEITAGI 240
QY 241 OKTARNDFLOSGMFKTRFDSLEYROTAKAAYELPEKDLPEEDFERARRELYOCFKRTDP 300
Db 241 OKTARNDFLOSGMFKTRFDSLEYROTAKAAYELPEKDLPEEDFERARRELYOCFKRTDP 300
QY 301 PSGAWLDFADGRMRRLFTLROYFGILYRLIKNRROARSDSAGKEDEI 348
Db 301 PSGAWLDFADGRMRRLFTLROYFGILYRLIKNRROARSDSAGKEDEI 348
```

RESULT 3

```
US-08-683-458-3
; Sequence 3, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683.458
; FILING DATE:
; CLASSIFICATION: 435
; PRIORITY APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-458-3
```

```
Query Match 100.0%; Score 1823; DB 1; Length 348;
Best Local Similarity 100.0%; Pred. No. 5.5e-190;
Matches 348; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

```
QY 1 MOPIVSVLICAVNVEKYFAQSILAAVYVNTWRNLDILIVDGSIDGTLAIAXDFQKRDSTRI 60
Db 1 MOPIVSVLICAVNVEKYFAQSILAAVYVNTWRNLDILIVDGSIDGTLAIAXDFQKRDSTRI 60
```


QY	61	KILQAOANSLGILPELNTGLDELASGGGGGYIARTADDIASPGWTEKIVGEMKDRSI	120
Db	61	KILQAOANSLGILPELNTGLDELASGGGGGYIARTADDIASPGWTEKIVGEMKDRSI	120
QY	121	IAMGAMLEVIUSEEKDGNRLAHNRHKGKIMWKPTRHEDIAPFPFGNDIHNTIMRASYI	180
Db	121	IAMGAMLEVIUSEEKDGNRLAHNRHKGKIMWKPTRHEDIAPFPFGNDIHNTIMRASYI	180
QY	181	DGGLRYPTERPMADYOFKWDVSKLGLAYYRELALVYRLAHNAVSSKHSVROHEIQAOGI	240
Db	181	DGGLRYPTERPMADYOFKWDVSKLGLAYYRELALVYRLAHNAVSSKHSVROHEIQAOGI	240
QY	241	OKTARNDFLQSMGKTRFSDLSLEYRQTKAAAYELPEKDLPEDEPERARFLEYOCFKRTDTP	300
Db	241	OKTARNDFLQSMGKTRFSDLSLEYRQTKAAAYELPEKDLPEDEPERARFLEYOCFKRTDTP	300
QY	301	PSGAMLPDPAADGRRRLFTLROYTGILYLRLIKNRQARSACKEDETI	348
Db	301	PSGAMLPDPAADGRRRLFTLROYTGILYLRLIKNRQARSACKEDETI	348

RESULT 4
US-08-878-360-3
; Sequence 3, Application US/08878360

```

1  GENERAL INFORMATION:
2  APPLICANT: Gotschlich, Emil C.
3  TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
4  TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
5  NUMBER OF SEQUENCES: 12
6  CORRESPONDENCE ADDRESS:
7  ADDRESSEE: Klauber & Jackson
8  STREET: 411 Hackensack Avenue
9  CITY: Hackensack
10 STATE: New Jersey
11 COUNTRY: USA
12 ZIP: 07601
13
14 COMPUTER READABLE FORM:
15 MEDIUM TYPE: Floppy disk
16 COMPUTER: IBM PC compatible
17 OPERATING SYSTEM: PC-DOS/MS-DOS
18 SOFTWARE: Patentin Release #1.0, Version #1.25
19
20 CURRENT APPLICATION DATA:
21 APPLICATION NUMBER: US/08/878,360
22 FILING DATE: 18-JUN-1997
23 CLASSIFICATION: 435
24
25 PRIOR APPLICATION DATA:
26 APPLICATION NUMBER: 08/683,426
27 FILING DATE:
28
29 APPLICATION NUMBER: 08/312,387
30 FILING DATE: September 26, 1994
31 CLASSIFICATION: 435
32
33 ATTORNEY/AGENT INFORMATION:
34 NAME: Jackson Esq., David A.
35 REGISTRATION NUMBER: 26,742
36 REFERENCE/DOCKET NUMBER: 600-1-095B
37
38 TELECOMMUNICATION INFORMATION:
39 TELEPHONE: 201 487-5800
40 TELEFAX: 201 343-1684
41
42 TELE: 133521
43
44 INFORMATION FOR SEQ ID NO: 3:
45 SEQUENCE CHARACTERISTICS:
46 LENGTH: 348 amino acids
47 TYPE: amino acid
48 TOPOLOGY: linear
49
50 MOLECULE TYPE: protein
51
52 US-08-878-360-3

```

Query Match	100.0%;	Score 1823;	DB 2;	Length 348;
Best Local Similarity	100.0%;	Pred. No. 5.5e-190;		
Matches 348;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

QY	1	MOPVSLICAVWEKVEYFPXOSLAAVWOTWNLIDLIVDSESTDTGLAIKDKQKXDSI	60
Db	1	MOPVSLICAVWEKVEYFPXOSLAAVWOTWNLIDLIVDSESTDTGLAIKDKQKXDSI	60
QY	61	KILAQAO NSGLIPSLNIGDELAKSGGGGGEXIARTADADIASPGWIEKIVGEMERDRI	12
Db	61	KILAQAO NSGLIPSLNIGDELAKSGGGGGEXIARTADADIASPGWIEKIVGEMERDRI	12
QY	121	IANGAMLEVLSEKDCGNRLARHHKHGKIWKPTRHEDIAAEFPFGNPIHNNTIMIRSVI	18
Db	121	IANGAMLEVLSEKDCGNRLARHHKHGKIWKPTRHEDIAAEFPFGNPIHNNTIMIRSVI	18
QY	181	DGGLRTDTERDMADYQFYWYVSKIGRLAATYREALVYKRLHANOVSSKSHVROHEITAGSI	24
Db	181	DGGLRTDTERDMADYQFYWYVSKIGRLAATYREALVYKRLHANOVSSKSHVROHEITAGSI	24
QY	241	OKTARNDFLQSMGFKRPPSLLEYROTKAAYVELPEKOLPREDFERRARFLYOCFKTTPR	30
Db	241	OKTARNDFLQSMGFKRPPSLLEYROTKAAYVELPEKOLPREDFERRARFLYOCFKTTPR	30
QY	301	PSGAMIDFADGEMRLFTPLROYFGILVRLIKNRQOASDAGKEDEI	348
Db	301	PSGAMIDFADGEMRLFTPLROYFGILVRLIKNRQOASDAGKEDEI	348

RESULT 5
US-09-333-412-3

Sequence ID: Application US/09333412
Patent No. 6342382
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: OLIGOSYSTRANSFERASES FOR BIOSYNTHESIS OF OLIGOSACCHARIDES, AND GENES ENCODING SAME
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
City: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: Linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-09-333-412-3

Query Match	100.0%;	Score 1823;	DB 4;	Length 348;
Best Local Similarity	100.0%;	Pred. No. 5.5e-190;		
Matches 348;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

```

01  KITAAQANSGSLTSLNIGIDELAKSGGGGGGYIARTDADIASPGWLEKTVGEMEKDRSI 12
02  DB 61 KITAAQANSGSLTSLNIGIDELAKSGGGGGGYIARTDADIASPGWLEKTVGEMEKDRSI 12
03  QY 121 IAMGAMLEVLSESEKDONRLARHHKKGKTKMKPTRHEDIAAFPPGNDTHNNTWIMRRSY 18
04  DB 121 IAMGAMLEVLSESEKDONRLARHHKKGKTKMKPTRHEDIAAFPPGNDTHNNTWIMRRSY 18
05  QY 181 DGLIRYDTERDMAEDYQFWYDVSKGLRLAYYPEALVKYRLHANOVSSKSHVROHEIAOGI 24
06  DB 181 DGLIRYDTERDMAEDYQFWYDVSKGLRLAYYPEALVKYRLHANOVSSKSHVROHEIAOGI 24
07  QY 241 QKTRANDFLQSMGFKTRRPSLEYRQTKAAAELEPKOLPEEDPRARRFYOCFKRTDTP 30
08  DB 241 QKTRANDFLQSMGFKTRRPSLEYRQTKAAAELEPKOLPEEDPRARRFYOCFKRTDTP 30
09  QY 301 PSGAMLDFADGMRRLFTLRQYFGILRLIKNRQARSADGKEQEI 348
10  DB 301 PSGAMLDFADGMRRLFTLRQYFGILRLIKNRQARSADGKEQEI 348
11  RESULT 7
12  US-08-683-426-11
13  : Sequence 11, Application US/08683426
14  : Patent No. 5705367
15  : GENERAL INFORMATION:
16  : APPLICANT: Gotschlich, Emil C.
17  : TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
18  : TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
19  : NUMBER OF SEQUENCES: 12
20  : CORRESPONDENCE ADDRESS:
21  : ADDRESSEE: Klauber & Jackson
22  : STREET: 411 Hackensack Avenue
23  : CITY: Hackensack
24  : STATE: New Jersey
25  : COUNTRY: USA
26  : ZIP: 07601
27  : COMPUTER READABLE FORM:
28  : MEDIUM TYPE: Floppy disk
29  : COMPUTER: IBM PC compatible
30  : OPERATING SYSTEM: PC-DOS/MS-DOS
31  : SOFTWARE: PatentIn Release #1.0, Version #1.25
32  : CURRENT APPLICATION DATA:
33  : APPLICATION NUMBER: US/08/683,426
34  : FILING DATE:
35  : CLASSIFICATION: 536
36  : PRIOR APPLICATION DATA:
37  : APPLICATION NUMBER: 08/312,387
38  : FILING DATE: September 26, 1994
39  : CLASSIFICATION: 536
40  : ATTORNEY/AGENT INFORMATION:
41  : NAME: Jackson Esq., David A.
42  : REGISTRATION NUMBER: 26,742
43  : REFERENCE/DOCKET NUMBER: 600-1-095B
44  : TELECOMMUNICATION INFORMATION:
45  : TELEPHONE: 201 487-5800
46  : TELEFAX: 201 343-1684
47  : TELEX: 133521
48  : INFORMATION FOR SEQ ID NO: 11:
49  : SEQUENCE CHARACTERISTICS:
50  : LENGTH: 348 amino acids
51  : TYPE: amino acid
52  : TOPOLOGY: linear
53  : MOLECULE TYPE: protein
54  : US-08-683-426-11
55  :
56  : Query Match 99.8% Score 1820; DB 1; Length 348;
57  : Best Local Similarity 99.7%; Pred. No. 12e-189;
58  : Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps:
59  :
60  : 1 MDPVSVLICAVNVEKYFAQSLAAVAVNQTWRNLDILIVDGSSTDGTALIAKDFQKRDSTI 60

```

Db 1 LQPLVSLICAVNVEKYFAQSLAAVYNQWTNRNDLIIYDGGSTDTGLTAKDFQKRDSTI 60
QY 61 KTLAQONSGILPSLNIIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
Db 61 KTLAQONSGILPSLNIIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
QY 121 IMGAWLEVLSEKOGNRLARHHKHKIWKKPTRHEDIAAFPEFGNPINNNTMIMRSYI 180
Db 121 IMGAWLEVLSEKOGNRLARHHKHKIWKKPTRHEDIAAFPEFGNPINNNTMIMRSYI 180
QY 181 DGLRYDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQGI 240
Db 181 DGLRYDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQGI 240
QY 241 OKTARNDFQSMGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRELYOCFKRTDTP 300
Db 241 OKTARNDFQSMGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRELYOCFKRTDTP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
Db 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348

RESULT 8

US-08-683-458-11
; Sequence 11, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-683-458-11

Query Match 99.8%; Score 1820; DB 1; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSLICAVNVEKYFAQSLAAVYNQWTNRNDLIIYDGGSTDTGLTAKDFQKRDSTI 60
Db 1 LQPLVSLICAVNVEKYFAQSLAAVYNQWTNRNDLIIYDGGSTDTGLTAKDFQKRDSTI 60
QY 61 KTLAQONSGILPSLNIIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
Db 61 KTLAQONSGILPSLNIIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEDRSI 120
QY 121 IMGAWLEVLSEKOGNRLARHHKHKIWKKPTRHEDIAAFPEFGNPINNNTMIMRSYI 180
Db 121 IMGAWLEVLSEKOGNRLARHHKHKIWKKPTRHEDIAAFPEFGNPINNNTMIMRSYI 180
QY 181 DGLRYDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQGI 240
Db 181 DGLRYDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQGI 240
QY 241 OKTARNDFQSMGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRELYOCFKRTDTP 300
Db 241 OKTARNDFQSMGFKTRFDSLEYRQTKAAAYELPEKDLPEEDFERARRELYOCFKRTDTP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
Db 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348

RESULT 9

US-08-878-360-11
; Sequence 11, Application US/08878360
; Patent No. 5945322
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,360
; FILING DATE: 18-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/683,426
; FILING DATE:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 11:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-878-360-11

Query Match 99.8%; Score 1820; DB 2; Length 348;

Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLICAVNEKVEKFAQSLAAVYVQWTRNDLILYDDGSTDGLTAAKDFOKRDSRI 60
DB 1 LOPVSVLICAVNEKVEKFAQSLAAVYVQWTRNDLILYDDGSTDGLTAAKDFOKRDSRI 60
QY 61 KILAAQONSGILPSNLIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
DB 61 KILAAQONSGILPSNLIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
QY 121 IAMGAWLEVLSEKDGKGNRLARHHKHKGIWKKPTRHEDIAAFPPGPNJHNTMIMRSTY 180
DB 121 IAMGAWLEVLSEKDGKGNRLARHHKHKGIWKKPTRHEDIAAFPPGPNJHNTMIMRSTY 180
QY 181 DGLRDTERRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLHANOVSSKSHVROHEIAOGI 240
DB 181 DGLRDTERRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLHANOVSSKSHVROHEIAOGI 240
QY 241 OKTARNDFLOSMGFKTRFDSLEYROTAKAAVELPEKDLPEEDPERARRFLYOCFKRTDP 300
DB 241 OKTARNDFLOSMGFKTRFDSLEYROTAKAAVELPEKDLPEEDPERARRFLYOCFKRTDP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
DB 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348

RESULT 10

US-08-478-140B-3
; Sequence 3, Application US/08478140B
; Patent No. 6127153

GENERAL INFORMATION:

APPLICANT: JOHNSON, KARL F.
APPLICANT: ROTH, STEPHEN
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,140B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-140B-3

Query Match 99.8%; Score 1820; DB 3; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLICAVNEKVEKFAQSLAAVYVQWTRNDLILYDDGSTDGLTAAKDFOKRDSRI 60
DB 1 LOPVSVLICAVNEKVEKFAQSLAAVYVQWTRNDLILYDDGSTDGLTAAKDFOKRDSRI 60
QY 61 KILAAQONSGILPSNLIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
DB 61 KILAAQONSGILPSNLIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
QY 121 IAMGAWLEVLSEKDGKGNRLARHHKHKGIWKKPTRHEDIAAFPPGPNJHNTMIMRSTY 180
DB 121 IAMGAWLEVLSEKDGKGNRLARHHKHKGIWKKPTRHEDIAAFPPGPNJHNTMIMRSTY 180
QY 181 DGLRDTERRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLHANOVSSKSHVROHEIAOGI 240
DB 181 DGLRDTERRDMAEDYQFMYDVSKLGRLAYYPEALVKYRLHANOVSSKSHVROHEIAOGI 240
QY 241 OKTARNDFLOSMGFKTRFDSLEYROTAKAAVELPEKDLPEEDPERARRFLYOCFKRTDP 300
DB 241 OKTARNDFLOSMGFKTRFDSLEYROTAKAAVELPEKDLPEEDPERARRFLYOCFKRTDP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
DB 301 PSGAWLDFADGMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348

RESULT 11

US-08-478-140B-8
; Sequence 8, Application US/08478140B
; Patent No. 6127153

GENERAL INFORMATION:

APPLICANT: JOHNSON, KARL F.
APPLICANT: ROTH, STEPHEN
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478,140B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
STRANDEDNESS: double
TOPOLOGY: linear
MOLECULE TYPE: protein

ANTI-SENSE: NO
US-08-478-140B-8

Query Match 99.8%; Score 1820; DB 3; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLICAVNVEKFFQSLAAVYVNTWRNLDLIYDGSSTDTGLATAKDFOKRDSRI 60
DB 1 LOPVSVLICAVNVEKFFQSLAAVYVNTWRNLDLIYDGSSTDTGLATAKDFOKRDSRI 60
QY 61 KILAAQNSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
DB 61 KILAAQNSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
QY 121 IAMGAMLEVLSEKDGKGNRLARHHKHKTIWKPTRHEDIAAFPPGPNHNMTMIMRSYI 180
DB 121 IAMGAMLEVLSEKDGKGNRLARHHKHKTIWKPTRHEDIAAFPPGPNHNMTMIMRSYI 180
QY 181 DGLRDTYTERDMAEDYQFVYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQSI 240
DB 181 DGLRDTYTERDMAEDYQFVYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQSI 240
QY 241 OKTARNDFLQSGFKTRFDSLEYROTAKAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
DB 241 OKTARNDFLQSGFKTRFDSLEYROTAKAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKEDEI 348
DB 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKEDEI 348

RESULT 12
US-09-333-412-11
Sequence 11, Application US/09333412
Patent No. 6342382
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear

MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-09-333-412-11

Query Match 99.8%; Score 1820; DB 4; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLICAVNVEKFFQSLAAVYVNTWRNLDLIYDGSSTDTGLATAKDFOKRDSRI 60
DB 1 LOPVSVLICAVNVEKFFQSLAAVYVNTWRNLDLIYDGSSTDTGLATAKDFOKRDSRI 60
QY 61 KILAAQNSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
DB 61 KILAAQNSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
QY 121 IAMGAMLEVLSEKDGKGNRLARHHKHKTIWKPTRHEDIAAFPPGPNHNMTMIMRSYI 180
DB 121 IAMGAMLEVLSEKDGKGNRLARHHKHKTIWKPTRHEDIAAFPPGPNHNMTMIMRSYI 180
QY 181 DGLRDTYTERDMAEDYQFVYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQSI 240
DB 181 DGLRDTYTERDMAEDYQFVYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVROHEIAQSI 240
QY 241 OKTARNDFLQSGFKTRFDSLEYROTAKAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
DB 241 OKTARNDFLQSGFKTRFDSLEYROTAKAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
QY 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKEDEI 348
DB 301 PSGAWLDFADGMRRLFTLRQYFGILYRLINRRQARSDSAGKEDEI 348

RESULT 13
US-09-338-943-3
Sequence 3, Application US/09338943
Patent No. 6379933
GENERAL INFORMATION:
APPLICANT: JOHNSON, KARL F.
APPLICANT: ROTH, STEPHEN
TITLE OF INVENTION: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/338,943
FILING DATE:
CLASSIFICATION:
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/478,140
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi
REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNIE

;; INFORMATION FOR SEQ ID NO: 3 :
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 348 amino acids
;; TYPE: amino acid
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
US-09-338-943-3

Query Match 99.8%; Score 1820; DB 4; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLPLSVLICAVNVEKYFQSLAAVYVNOTWRMDILIVDGSSTDTGLTAAKDFOKRDSRI 60
:|||||
DB 1 LQPLVSVLICAVNVEKYFQSLAAVYVNOTWRMDILIVDGSSTDTGLTAAKDFOKRDSRI 60
QY 61 KILAAQONSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
:|||||
DB 61 KILAAQONSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
QY 121 IAMGAWLEYLESEKGNRLARHHKHGKIWKKPTRHEDIAAFPPGNPIHNTMTIMRRSYI 180
:|||||
DB 121 IAMGAWLEYLESEKGNRLARHHKHGKIWKKPTRHEDIAAFPPGNPIHNTMTIMRRSYI 180
QY 181 DGLGRYDPERDMAEDYQFYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVRQHEIAGSI 240
:|||||
DB 181 DGLGRYDPERDMAEDYQFYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVRQHEIAGSI 240
QY 241 OKTARNDFLOSGMFKTRPDSLEYRQTKAAAYELPEKDLPEEDFEERARRELYOCFKRTDTP 300
:|||||
DB 241 OKTARNDFLOSGMFKTRPDSLEYRQTKAAAYELPEKDLPEEDFEERARRELYOCFKRTDTP 300
QY 301 PSGAWLDFPADGRMRRLFTLRQYFGILYRLINRRQARSDSGKDEOI 348
:|||||
DB 301 PSGAWLDFPADGRMRRLFTLRQYFGILYRLINRRQARSDSGKDEOI 348

RESULT 14

US-09-338-943-8
; Sequence 8, Application US/09338943
; Patent No. 6379933
; GENERAL INFORMATION:

;; APPLICANT: JOHNSON, KARL F.
;; APPLICANT: ROTH, STEPHEN
;; APPLICANT: BOCCAZZA, STEPHANIE L.
;; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
;; TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
;; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
;; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
;; NUMBER OF SEQUENCES: 8
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Pennie & Edmonds
;; STREET: 1155 Avenue of the Americas
;; CITY: New York
;; STATE: New York
;; COUNTRY: USA
;; ZIP: 10036-2711

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.30
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/09/338,943
;; FILING DATE:
;; CLASSIFICATION:
;; PRIOR APPLICATION DATA:
;; APPLICATION NUMBER: US 08/478,140
;; FILING DATE:
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Laura A. Coruzzi
;; REGISTRATION NUMBER: 30,742
;; REFERENCE/DOCKET NUMBER: 7188-017

;; TELECOMMUNICATION INFORMATION:
;; TELEPHONE: (212) 790-9090
;; TELEFAX: (212) 869-9741/8864
;; TELEX: 66141 PENNIE
;; INFORMATION FOR SEQ ID NO: 8 :
;; SEQUENCE CHARACTERISTICS:
;; LENGTH: 348 amino acids
;; TYPE: amino acid
;; STRANDEDNESS: double
;; TOPOLOGY: linear
;; MOLECULE TYPE: protein
;; AMTI-SENSE: NO
US-09-338-943-8

Query Match 99.8%; Score 1820; DB 4; Length 348;
Best Local Similarity 99.7%; Pred. No. 1.2e-189;
Matches 347; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLPLSVLICAVNVEKYFQSLAAVYVNOTWRMDILIVDGSSTDTGLTAAKDFOKRDSRI 60
:|||||
DB 1 LQPLVSVLICAVNVEKYFQSLAAVYVNOTWRMDILIVDGSSTDTGLTAAKDFOKRDSRI 60
QY 61 KILAAQONSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
:|||||
DB 61 KILAAQONSGLIPSLINIGDELAKSGGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
QY 121 IAMGAWLEYLESEKGNRLARHHKHGKIWKKPTRHEDIAAFPPGNPIHNTMTIMRRSYI 180
:|||||
DB 121 IAMGAWLEYLESEKGNRLARHHKHGKIWKKPTRHEDIAAFPPGNPIHNTMTIMRRSYI 180
QY 181 DGLGRYDPERDMAEDYQFYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVRQHEIAGSI 240
:|||||
DB 181 DGLGRYDPERDMAEDYQFYDVSKLGRLAYYPEALVKYRLHANOVSSKHSVRQHEIAGSI 240
QY 241 OKTARNDFLOSGMFKTRPDSLEYRQTKAAAYELPEKDLPEEDFEERARRELYOCFKRTDTP 300
:|||||
DB 241 OKTARNDFLOSGMFKTRPDSLEYRQTKAAAYELPEKDLPEEDFEERARRELYOCFKRTDTP 300
QY 301 PSGAWLDFPADGRMRRLFTLRQYFGILYRLINRRQARSDSGKDEOI 348
:|||||
DB 301 PSGAWLDFPADGRMRRLFTLRQYFGILYRLINRRQARSDSGKDEOI 348

RESULT 15

US-08-312-387B-5
; Sequence 5, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:

;; APPLICANT: Gotschlich, Emil C.
;; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
;; NUMBER OF SEQUENCES: 12
;; CORRESPONDENCE ADDRESS:
;; ADDRESSEE: Klauber & Jackson
;; STREET: 411 Hackensack Avenue
;; CITY: Hackensack
;; STATE: New Jersey
;; COUNTRY: USA
;; ZIP: 07601

;; COMPUTER READABLE FORM:
;; MEDIUM TYPE: Floppy disk
;; COMPUTER: IBM PC compatible
;; OPERATING SYSTEM: PC-DOS/MS-DOS
;; SOFTWARE: Patentin Release #1.0, Version #1.25
;; CURRENT APPLICATION DATA:
;; APPLICATION NUMBER: US/08/312,387B
;; FILING DATE: July 7, 1994
;; CLASSIFICATION: 435
;; ATTORNEY/AGENT INFORMATION:
;; NAME: Jackson Esq., David A.
;; REGISTRATION NUMBER: 26,742
;; REFERENCE/DOCKET NUMBER: 600-1-095
;; TELECOMMUNICATION INFORMATION:

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 12:00:15 : Search time 6.97495 Seconds
(without alignments)
794.504 Million cell updates/sec

Title: US-10-007-267-3
Perfect score: 1823
Sequence: 1 MQLPVSYLCAVNEYKFAQ.....RLIKNRQARSASAKQEDEI 348

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PC1_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/PC1US_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEW_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1823	100.0	348	12	US-10-007-267-3
2	1820	99.8	348	12	US-10-007-267-11
3	1049	57.5	337	12	US-10-007-267-5
4	1046	57.4	337	12	US-10-007-267-12
5	230	12.6	150	10	US-09-924-358-29
6	228.5	12.5	332	10	US-09-767-041-22
7	227	12.5	278	10	US-09-767-041-36
8	210.5	11.5	313	10	US-09-900-038A-1
9	210	11.5	322	10	US-09-767-041-34
10	208.5	11.4	972	9	US-09-879-959-10
11	195	10.7	301	10	US-09-816-028A-27
12	194.5	10.7	322	10	US-09-767-041-35
13	194	10.6	303	10	US-09-816-028A-29
14	191.5	10.5	674	10	US-09-765-272-200
15	190	10.4	270	10	US-09-816-028A-39
16	183	10.0	332	10	US-09-767-041-21
17	179	9.8	389	10	US-09-816-028A-34
18	176	9.7	120	10	US-09-767-041-52
19	175	9.6	270	10	US-09-765-272-198

20	175	9.6	281	10	US-09-765-272-196	Sequence 196, App
21	169	9.3	297	10	US-09-816-028A-31	Sequence 31, Appl
22	168.5	9.2	120	10	US-09-767-041-51	Sequence 51, Appl
23	164	9.0	706	10	US-09-815-242-1950	Sequence 4850, Ap
24	164	9.0	710	10	US-09-815-242-10511	Sequence 10511, A
25	158	8.7	715	10	US-09-767-041-47	Sequence 47, Appl
26	151	8.3	187	9	US-09-973-457-4	Sequence 4, Appl
27	151	8.3	187	10	US-09-815-028-7	Sequence 7, Appl
28	151	8.3	187	12	US-10-074-527-4	Sequence 4, Appl
29	147.5	8.1	79	10	US-09-765-272-168	Sequence 168, App
30	142.5	7.8	269	10	US-09-767-041-41	Sequence 41, Appl
31	129	7.1	303	10	US-09-765-272-202	Sequence 202, App
32	123	6.7	358	10	US-09-815-242-5714	Sequence 5714, Ap
33	123	6.7	573	10	US-09-815-242-12474	Sequence 12474, A
34	116.5	6.4	559	12	US-10-001-851-22	Sequence 22, Appl
35	113.5	6.2	559	12	US-10-001-851-20	Sequence 20, Appl
36	113.5	6.2	559	12	US-10-001-851-21	Sequence 21, Appl
37	113.5	6.2	559	12	US-10-001-851-23	Sequence 23, Appl
38	113.5	6.2	561	10	US-09-925-301-1006	Sequence 1006, Ap
39	111.5	6.1	612	12	US-10-001-851-25	Sequence 25, Appl
40	110.5	6.1	559	12	US-10-001-851-24	Sequence 24, Appl
41	104	5.7	418	10	US-09-816-028A-33	Sequence 33, Appl
42	102	5.6	626	12	US-10-001-851-27	Sequence 27, Appl
43	100.5	5.5	608	10	US-09-924-358-8	Sequence 8, Appl
44	98.5	5.4	256	10	US-09-925-301-883	Sequence 883, App
45	97.5	5.3	326	9	US-10-108-605-21	Sequence 21, Appl

ALIGNMENTS

RESULT 1
US-10-007-267-3
Sequence 3, Application US/10007267
Patent No. US20020127682A1
GENERAL INFORMATION:
APPLICANT: Gottschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid

TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-007-267-5

Query Match 57.5%; Score 1049; DB 12; Length 337;
Best Local Similarity 64.5%; Pred. No. 7.1e-91;
Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

QY 1 MOPLVSVLICAVNVEKYFAQSLAAVAVNQTWRNLDILYDDGSTDGTALAKDFQRDSRI 60
DB 1 MOPLVSVLICAVNAKRYFAQSLAAVAVGQTWRNLDILYDDGSTDGTALAKHFGQDGR 60
QY 61 KILAAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
DB 61 RLISNPRNLGFIASLNGIDELAKS--GGGEYIARTDADDIASPGMIKIVGEMEKDRSI 118
QY 121 IAMGAMLEYLESEKDNRLARHKGKTKKPTRHEDIAAFPPGPNPHNNTMIRRSYI 180
DB 119 IAMGAMLEYLESENNKSVLAATARNGAIWDKPTRHEDIAAFPPGPNPHNNTMIRRSYI 178
QY 181 DGLGAYTERDWAEDYQFVYDKLGRALAYYPEALVKYRLAHANOVSSKSHVROHEIAQI 240
DB 179 DGLGFDPAYIAEDYKFWYKGLGRALAYYPEALVKYRFHODQSSKYNLQORRTAMKI 238
QY 241 OKTARNDFLOSNGFTRPSLEYROTKAAYELPEKDLPEEDFEARARRELYCCFRRTDP 300
DB 239 KEIRIAGYWKAGIAGVADCLNLYGLKSTAYALYERKALSGDIGCLRFLYFLSLEY 298
QY 301 PSGAMLDFAADGRMRFLTRQYFGLYRLIK 332
DB 299 SLTDLLDPLTDPRVMRKLFAPQYRKILKKMLR 330

RESULT 4

US-10-007-267-12
Sequence 12, Application US/10007267
Patent No. US20020127682A1
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
OPERATING SYSTEM: IBM PC compatible
SOFTWARE: Patent Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
APPLICATION NUMBER: 08/312,387
FILING DATE: JULY 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 12:
SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-007-267-12

Query Match 57.4%; Score 1046; DB 12; Length 337;
Best Local Similarity 64.2%; Pred. No. 1.4e-90;
Matches 213; Conservative 34; Mismatches 83; Indels 2; Gaps 1;

QY 1 MOPLVSVLICAVNVEKYFAQSLAAVAVNQTWRNLDILYDDGSTDGTALAKDFQRDSRI 60
DB 1 MOPLVSVLICAVNAKRYFAQSLAAVAVGQTWRNLDILYDDGSTDGTALAKHFGQDGR 60
QY 61 KILAAQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
DB 61 RLISNPRNLGFIASLNGIDELAKS--GGGEYIARTDADDIASPGMIKIVGEMEKDRSI 118
QY 121 IAMGAMLEYLESEKDNRLARHKGKTKKPTRHEDIAAFPPGPNPHNNTMIRRSYI 180
DB 119 IAMGAMLEYLESENNKSVLAATARNGAIWDKPTRHEDIAAFPPGPNPHNNTMIRRSYI 178
QY 181 DGLGAYTERDWAEDYQFVYDKLGRALAYYPEALVKYRLAHANOVSSKSHVROHEIAQI 240
DB 179 DGLGFDPAYIAEDYKFWYKGLGRALAYYPEALVKYRFHODQSSKYNLQORRTAMKI 238
QY 241 OKTARNDFLOSNGFTRPSLEYROTKAAYELPEKDLPEEDFEARARRELYCCFRRTDP 300
DB 239 KEIRIAGYWKAGIAGVADCLNLYGLKSTAYALYERKALSGDIGCLRFLYFLSLEY 298
QY 301 PSGAMLDFAADGRMRFLTRQYFGLYRLIK 332
DB 299 SLTDLLDPLTDPRVMRKLFAPQYRKILKKMLR 330

RESULT 5

US-09-924-358-29
Sequence 29, Application US/09924358
Patent No. US20020107376A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Meyers, Rachel
TITLE OF INVENTION: 26199, 33530, 33949, 47148, 50226, AND
TITLE OF INVENTION: 58764,
FILE REFERENCE: 38155-20034..00
CURRENT APPLICATION NUMBER: US/09/924,358
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/229,300
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 150
TYPE: prt
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus amino acid
US-09-924-358-29

Query Match 12.6%; Score 230; DB 10; Length 150;
Best Local Similarity 40.8%; Pred. No. 1.5e-14;
Matches 53; Conservative 29; Mismatches 38; Indels 10; Gaps 5;

	Query Match	12.5%	Score 228.5;	DB 10;	Length 332;	
	Best Local Similarity	23.6%;	Pred. No. 6.1e-14;			
	Matches	87;	Conservative	68;	Mismatches 99;	Indels 115; Gaps 16
Oy	5	VSLICAYNVEKYFAQOSLAAYVNYQTNRNDILLIVDDSGTDTGLATAKDEQOKRDSRIKIILA	64			
	:	:::: :::: :				
Db	4	ISIIPTYNVEGELSKINSIYNGTYKHITIELLVNDGSIDNSEGEITLAVAKKDSRIYRF-	62			
Oy	65	QAQNSGLPIRLNTIGDELAKSGGGGGYATRPDADDIASPGMIEIKTVGEMEKDRSIAMG	124			
	:	: :				
Db	63	KKEKGSLDARUYGISLR-AK-----GDYLAFIDSDFEISEFIQRLEHAEIERENALVAVA	116			
Oy	125	AWLEYLSEKDGONRLAHNNHKGIKKKPTRHEDIAAFEPFGAPRIHNNTIMR-RSVIDGG	183			
	:	: :				
Db	117	GIDRY-----DASGHFLTAERLPINQVALSRNCKKL	149			
Oy	184	LRYDTER---DW-----AEDYOFWMD-VSKIGRLAAYPEALVYKRLH	221			
	:	: :				
Db	150	LEADGRHFVAANKLYKKELFDFFREKGIHDEDEPTYRLYLELEKVALVKECLYYYVD	209			
Oy	222	ANQV--SKSHVRQHEIDAGIOKTARNPDLQSMGFRTSDLSLEYRKTKAAAYELPRKDL	279			
	:	: :				
Db	210	ENSIIITSMTDHRFCLLE-FONERIMDYESRG-----DKEL-	245			
Oy	280	EEDFERARRFLIQCKRTDPSPGAMLDEPA-----DGRMRLLFTLRQYFGILVR	329			
	:	: :				
Db	246	-----LLECIR-----SFLAFAYLFEGKNYMWLSKQCKKLTLL--FRIVYK	284			
Oy	330	LTK-NRRQA	337			
	:	: :				
Db	285	OLKONKRILA	293			

```

RESULT 8
US-09-900-038A-1
; Sequence 1, Application US/09900038A
; Patent No. US20020142425A1
; GENERAL INFORMATION:
; APPLICANT: Miyake, Katsuhide
; APPLICANT: Watanabe, Masaki
; APPLICANT: Iijima, Shinji
; TITLE OF INVENTION: Beta 1,3-galactosyltransferase and DNA encoding the same
; FILE REFERENCE: 766.53
; CURRENT APPLICATION NUMBER: US/09/900,038A
; CURRENT FILING DATE: 2001-09-21
; PRIOR APPLICATION NUMBER: JP 2001-392
; PRIOR FILING DATE: 2001-01-05
; NUMBER OF SEQ ID NOS: 8
; SOFTWARE: Patentlin Ver. 2.0
; SEQ ID NO 1
; LENGTH: 313
; TYPE: PRT
; ORGANISM: Streptococcus agalactiae Type Ib
US-09-900-038A-1

```

Query Match	11.5%;	Score 210.5;	DB 10;	Length 313;
Best Local Similarity	27.8%;	Pred. No. 2.7e-12;		
Matches 86;	Conservative 55;	Mismatches 123;	Indels 45;	Gaps 15

```

QY      6  SVLLCAVNE-KYPAQSLAAVNOTMNLILLIYDDCSSTOYL-AIANKPOKRSRKIL 63
Db      4  SIMSVNEPLANTYRDVSIESLNOTLUDFEELIYIDNPSRDLKQFLFEYSVDNRKIL 63
QY      64  AQAQNSGLIPSLNGLDELAKSGGSGGEYIARTDADIAPSGWTEKIVGEMEKDRSIAM 123
Db      64  LNEENIGLASSLNAAV-KISK----GEYIFRMADDIAPSRDKOIRMEENS--LDF 115
QY      124  GAWLEVLSEEDGNRLARHNHKGKIKRKPTREHEDIAEPFGPNIHNNMTIMRSYIDGG 183
Db      116  SATLEILIDQK-GULVYVKQRESNKIYTL---NIRKMLNRSILAPTCOVKKKVEDKL 170
QY      184  LRYTERDW--AEYQWYDYSKIGRLAYP----EALVKYRLHANQVSKSHSVROH-- 23
Db      171  MGY---RDLPVEVEDYD---AIRGALADFIIGLLNVLLOYLRINENGISQTNFKQYIY 223
QY      235  -EIAQGIQK-----TARNDFLOSMGFKTRFDSLEKRYQTKAAAYELPEKDLPEDEERA 286
Db      224  SALQDFEYKESYIDITRKITNYFQEVYIKKNYT---QQLSKVFEL--KSTPSITIRKL 277
QY      287  RRFIYQCFK 295
Db      278  YICLYLYFK 286

```

RESULT 9
US-09-767-041-34
; Sequence 34, Application US/09767041

```

: GENERAL INFORMATION:
: APPLICANT: Smith, Hilda
: TITLE OF INVENTION: STREPTOCOCCUS SUIS VACCINES AND DIAGNOSTIC TESTS
: FILE REFERENCE: 2183-4776
: CURRENT APPLICATION NUMBER: US/09/767,041
: CURRENT FILING DATE: 2001-01-22
: PRIOR APPLICATION NUMBER: PCT/NL99/00460
: PRIOR FILING DATE: 1999-07-19
: PRIOR APPLICATION NUMBER: EP99202465.5
: PRIOR FILING DATE: 1998-07-22
: PRIOR APPLICATION NUMBER: EP99202467.1
: PRIOR FILING DATE: 1998-07-22
: NUMBER OF SEQ ID NOS: 53
: SOFTWARE: PatentIn version 3.0
: SEQ ID NO 34
: LENGTH: 322
: TYPE: prt
: ORGANISM: Streptococcus suis
: FEATURE:
: NAME/KEY: misc_feature
: OTHER INFORMATION: CPSII
: US-09-767-041-34

```

Query Match	11.5%	Score 210	DB 10	Length 322
Best Local Similarity	23.1%	Pred. No. 3.2e-12		
Matches 74	Conservative 66	Mismatches 124	Indels 56	Gaps 12

[illegible]

Qy	179	VIDGRLRDETRDAEAYQVFWYDV-SKLGRA-----	YEEPAALKYRLHANVSSKHSVR	232
Db	167	IKDKRQINRRSGCEDLFLVELLNWTVRVVDTRREYYNVIR--	NSSLINQPSIN	223
Qy	233	QHETIAQIQKTAARNDFLOSQMGKTRPDSLEYRQKKAAYE-----	LP	274
Db	224	NIDLTVR-----LENYFPRKAKREFSHYFDPAKVIKEKYCLNKMYSTDCLDNEFLPI		274
Qy	275	-----EKDLPEDEFERARRFL	290	
Db	275	LESTRKEIRRRYPTKAKRYL	294	

```

RESULT 10
US-09-879-959-10
: Sequence 10, Application US/09879959
: Patent No. US20020160489A1
:
: GENERAL INFORMATION:
:
: APPLICANT: Weigel, Paul H
: APPLICANT: Kumari, Kshama
: APPLICANT: Deangelis, Paul
:
: TITLE OF INVENTION: STREPTOCOCCUS EQUISIMILIS HYALURONAN SYNTHASE GENE AND EXPRESSION
:
: TITLE OF INVENTION: IN BACILLUS SUBTILIS
:
: FILE REFERENCE: 3554.049
:
: CURRENT APPLICATION NUMBER: US/09/879,959
:
: CURRENT FILING DATE: 2001-09-12
:
: PRIOR APPLICATION NUMBER: 09/469,200
:
: PRIOR FILING DATE: 1999-12-21
:
: PRIOR APPLICATION NUMBER: 09/178,851
:
: PRIOR FILING DATE: 1998-10-26
:
: NUMBER OF SEQ ID NOS: 10
:
: SOFTWARE: PatentIn version 3.1
:
: SEQ ID NO 10
:
: LENGTH: 972
:
: TYPE: PRT
:
: ORGANISM: pasteurella multocida
US-09-879-959-10

```

Query Match	11.4%	Score 208.5	DB 9	Length 972
Best Local Similarity	26.2%	Pred. No. 2e-11		
Matches 75	Conservative 56	Mismatches 108	Indels 47	Gaps 13

QY	3	PLVSLVCAYVEXEYFPOSLAAVYVOWJRNNDILIVDGGSDGLTALAKOFKRDSTIKI	62
		: : : : : :	
Db	440	PLVSLIPAYKAMAYIQRCVDSALNQIVVDLEVTCIDGSDTDTLEVITANKLYGNPNVRI	499
QY	63	LAQONSGILPSLNIGLDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKRSIIA	122
		:: : : : : : : :	
Db	500	MSK--PNGGSIASASNAAV--SFAK-----GYIYQGLDSDLEPDDVETCLKEPLKDKTLAC	552
QY	123	MGWILEVLSEKDCGNRLRAHHKHCITKKPRHEDIAAFEPFGAPINNNMIMRS--VI	180
		:: : : : : : : : : : : : : : :	
Db	553	V--YTTNNVNPDSLSLA---NMYNMPFSREKLTLTAMT-----AHHPMPFTIRAHHLT	601
QY	181	DGGLRYPDTERDMAEDYOFWYDVSXKIGRLAYAYPEALVYVYRIHANOVSSKSHSRQHEIAGOI	240
		: : : : : : : : : : : : : : :	
Db	602	DG-----FNKEIENAVDYDMFLKLSVGFKHLNLCYNRVLHGDNWTSIK-----KLG	650
QY	241	OKTARNDFL---QSMGR-----TFPDSL-----YRQTKAAYE	272
		: : : : : : : : : : : : : : :	
Db	651	QK--KNHEVVVNOSLNRGITYVNYVDEFFDDLSRKYIFNKTAQY	694

RESULT 11
 US-09-816-028A-27
 : Sequence 27, Application US/09816028A
 : Patent No. US20020042369A1
 : GENERAL INFORMATION:
 : APPLICANT: Gilbert, Michel
 : APPLICANT: Wakarchuk, Warren W.
 : APPLICANT: National Research Council of Canada
 : TITLE OF INVENTION: Camphylodacer Glycosyltransferases for Biosynthesis of
 : TITLE OF INVENTION: Gangliosides and Ganglioside Mimics

```
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 27
; LENGTH: 301
; TYPE: PR
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: beta-1,3-galactosyltransferase from C. jejuni strain
; OTHER INFORMATION: OH4384 (ORF 6a of 11pooligosaccharide (LOS))
; OTHER INFORMATION: biosynthesis locus)
US-09-816-028A-27
```

```
Query Match          10.7%; Score 195; DB 10; Length 301;
Best Local Similarity 24.4%; Pred. No. 7,3e-11;
Matches 71; Conservative 66; Mismatches 100; Indels 54; Gaps 12;
```

```
OY 5 VSVLICAVNVEKYFAOSLAAVNQTWRNLDLIVDGSFTGTLAAKDFQKRSRIKILA 64
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 4 ISILPTVNEQYIARIESCINOTFKDIELIYVDCGNDNSINAIKEYSKKDKRIKIH 63

OY 65 QAOSSGLIPSLNIGLDELASGGGGEGYIARTDADIASPGWIE---KIYGEKMDRSII 121
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 64 NEKMGLIKRAREYGV-KVANS-----PYIMFLDDPYLELNACECIKILDEED-EVDLV 116

OY 122 AMGAMLE--VLSEER-DGNR-----LARRHKGKIMKPTRRHEDIAAFPEFGN 166
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 117 FFMATIVESNVISYKKFEDNSGFYSKKEFVKIIAKKMLVMTMCK----- 161

OY 167 PIIHNTIMKRSYIDG---GLRYTERDAEDYQFWYD-VSKLGLAYPEALVYRLH 221
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 162 -----IIRKKYLEAFASLRLEKDVKINMAEDVLVYPMISQAKTAYMNCNLHYVPN 215

OY 222 ANOV-SSKHSVRQHEIAOGIOKTARNDFLOSMGFKTRFDSLEYQTAAY 271
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 216 NNSICMKNREYLVKNNIQELQVL--NYLKNQYILNKYCSVLVLYLKYLLY 264
```

```
RESULT 12
US-09-767-041-35
; Sequence 35, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP98202465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP98202467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 53
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO 35
; LENGTH: 322
; TYPE: PR
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc.feature
; OTHER INFORMATION: CPSIJ
US-09-767-041-35
```

```
Query Match          10.7%; Score 194.5; DB 10; Length 322;
Best Local Similarity 24.7%; Pred. No. 9e-11;
```

```
Matches 76; Conservative 44; Mismatches 97; Indels 91; Gaps 14;

OY 5 VSVLICAVNVEKYFAOSLAAVNQTWRNLDLIVDGSFTGTLAAKDFQKRSRIKILA 64
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 4 ISVIVPYNDVKIUSCIESIIMONKNIETLILDDSVDSKICKYEYK-DKRAVIF 62

OY 65 QAOSSGLIPSLNIGLDELASGGGGEGYIARTDADIASPGWIEKIYGEKMDRSII 124
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 63 -THSGVSNARNNGIKR-----STAEYIMFVSDVDVDSRLVEKLYFNKRSRDS-G 114

OY 125 AMLEVLSEKDCNRRLARRHKGKIMKPTRRHEDIAAFPEFGNPPIHNTIMRSYIDG- 183
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 115 CLYATFSEENINN-----FEVNN--NIDFEAINTVDDMGE 147

OY 184 -----LRYDT---ERDW-AEDYGF-WYDYSKGLRLAYPEALV 216
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 148 KFMNLYINNIESTPYCKLYKRIITDLEFENQWLGEDLLFNLYLKNIDRVSTLEHLY 207

OY 217 KYRLHANQVSSKHSVRQHEIAOGIOKTARNDFLOSMGFKTRFDSLEYQTKAAAYELPER 276
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 208 FYR-----RGLSTFV-NSFKRGV-----FQLENLQKQVLYL---FK 240

OY 277 DLPEDEFE 284
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 241 QIYGEDFD 248
```

```
RESULT 13
US-09-816-028A-29
; Sequence 29, Application US/09816028A
; Patent No. US20020042369A1
; GENERAL INFORMATION:
; APPLICANT: Wakarchuk, Michel
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter glycosyltransferases for Biosynthesis of
; FILE REFERENCE: 019633-000111US
; CURRENT APPLICATION NUMBER: US/09/816,028A
; CURRENT FILING DATE: 2001-03-21
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 29
; LENGTH: 303
; TYPE: PR
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: Campylobacter glycosyltransferase B (CgtB) beta-1,3
; OTHER INFORMATION: galactosyltransferase from C. jejuni serotype O:2
; OTHER INFORMATION: (strain NCTC 11168)
US-09-816-028A-29
```

```
Query Match          10.6%; Score 194; DB 10; Length 303;
Best Local Similarity 23.8%; Pred. No. 9,2e-11;
Matches 62; Conservative 66; Mismatches 97; Indels 36; Gaps 9;

OY 5 VSVLICAVNVEKYFAOSLAAVNQTWRNLDLIVDGSFTGTLAAKDFQKRSRIKILA 64
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 4 ISILPTVNEQYIARIESCINOTFKDIELIYVDCGNDNSINAIKEYSKDRIRKIH 63

OY 65 QAOSSGLIPSLNIGLDELASGGGGEGYIARTDADIASPGWIEKIYGEKMDRSII 124
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 64 NEEMLKILRAREYGA-KVANS-----PYIMFLDSDPYLELNACECI-----KILDMG 110

OY 125 AMLE-----VLSEKGNRLARRHKGKIMKPTRRHEDIAAFPEF---GNPIHNN 171
:|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 111 GGGKIDLLCFEAFITTNKAKSIKYL--NIKOGKYNNKEFTMOLIKTNPFMTMAKIKIKD 168

OY 172 TMIKRSVIDGLRYTERDAEDYQFWYDVKL-GLAYPEALVYKRLHANOV----- 225
```

Db 169 IYKARNMN--LKKKIKINMADALLYPLTILNSIEFYLTQPLYTOHNSNSTNNIN 226
OY 226 SSKSHVROHEIAOGIOKTARN 246
Db 227 SLEANTOEHKIVLNLKSTKN 247

RESULT 14
US-09-765-272-200

; Sequence 200, Application US/09765272
; Patent No. US20020061545A1
; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigens and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/765,272
; FILING DATE: 22-Jan-2001
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/961,083
; FILING DATE: <Unknown>
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: 36,373
; REFERENCE/DOCKET NUMBER: PB340P2
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8512
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 200:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 674 amino acids
; TYPE: amino acid
; STRANDEDNESS: single
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; SEQUENCE DESCRIPTION: SEQ ID NO: 200:
US-09-765-272-200

Query Match 10.5%; Score 191.5; DB 10; Length 674;
Best Local Similarity 21.2%; Pred. No. 4.7e-10;
Matches 68; Conservative 64; Mismatches 102; Indels 87; Gaps 13;

OY 22 LAAVVOTNRNLDILIVDGSSTGTLAKDPKRSRIKTLAQAONSGILPSLNLGIDE 81
Db 2 LDSITITQTNITIVVNDOSDASEICEFSEMDRRIITYIQ-ENAGISAARNGTGLN 60
OY 82 LAKSGGGGGEYIARTADIASPGWIEKIVGEMEKRSIITAMGAMLEVLSEKDCGNLAR 141
Db 61 MS-----GNYVFEVSDMIEDQYVELTKKIVEXQADIAVGNYSF--NESEG--MEY 110
OY 142 HHKHGKIMKPTTHEDIAAFPPGPNIHNTMIMRSVIDGCLRYDTER-----DWA 193
Db 111 FHILGD-----SYEKVYDVNSIFEN-----LYETQEMKSFALISAMG 148
OY 194 EDY-----QFWYDVSIG-----RLAYPEALVKYRLHANQVSSKHSVR- 232
Db 149 KLYKALFEOJLRDGLKEDGILNOKVYLSKRVLYLNKSLAYRIKRGSLSRVWTEKW 208
OY 233 OHEIAOGIOKTARNDFLOSNGEFTKRFDSLEYRO-----TKAAAYELPEKDL 278

Db 209 MHALVAMSE--RTITLLANGYPLEKHLAVYROMLEVSLANQASGLSPATRY----- 259
OY 279 PEEDFERARRELYQCFKRTDT 299
Db 260 --KEFEKORLLNQLSRQES 278

RESULT 15
US-09-816-028A-39

; Sequence 39, Application US/09816028A
; Patent No. US20020042369A1
; GENERAL INFORMATION:
; APPLICANT: Gilbert, Michel
; APPLICANT: Wakarchuk, Warren W.
; APPLICANT: National Research Council of Canada
; TITLE OF INVENTION: Campylobacter glycosyltransferases for Biosynthesis of
; FILE REFERENCE: 019633-00011105
; CURRENT APPLICATION NUMBER: US/09/816,028A
; PRIOR APPLICATION NUMBER: US 60/118,213
; PRIOR FILING DATE: 1999-02-01
; PRIOR APPLICATION NUMBER: US 09/495,406
; PRIOR FILING DATE: 2000-01-31
; NUMBER OF SEQ ID NOS: 49
; SOFTWARE: PatentIn Ver. 2.1
; SEQ ID NO 39
; LENGTH: 270
; TYPE: PRT
; ORGANISM: Campylobacter jejuni
; FEATURE:
; OTHER INFORMATION: glycosyltransferase from C. jejuni OH4384 (ORF 12a
; US-09-816-028A-39

Query Match 10.4%; Score 190; DB 10; Length 270;
Best Local Similarity 22.0%; Pred. No. 1.9e-10;
Matches 66; Conservative 54; Mismatches 102; Indels 78; Gaps 9;

OY 3 PLVSLICAVNVEKFRASLAAYVNOTWRNLDILIVDGSSTGTLAIADFOKRSRIKI 62
Db 2 POLSITIPFNSCDFISRALQSCINOTLKDIETILIDDKSKDNLNMYLEFAKKDPRIKI 61
OY 63 LAQAONSGILPSLNLGIDELAKSGGGGEYIARTADIASPGWIEKIVGEMEKDSIILA 122
Db 62 FQNEENLGTFFASRNLDVLA-----SSDFTMFLDSDDFLTPDACEIAFEKMKGFDLIC 115
OY 123 MGA-----WLEVLSEKDCGNRLARHHKHKGIWKKPTRHEDIA 159
Db 116 FDAFVHRVKTQFYRKQDEYFQKKEFLFLSKOR-----HFCMSVAKCFKKDIL 167
OY 160 APEPPGNPIHNTMIMRSVIDGCLRYDTERDWAEDYQFWYDYSKI--GRLAYVPEALVKY 218
Db 168 KSPE-----KIKIDERLNYGCEVLFYIYFMFCEKIAVFKTCIYHY 208
OY 219 RLHAN---QVSSKHSVRQ-----HEIAOGIOKT-ARNDPLQSM-----GFKTR 257
Db 209 EEPNNGRYENKKEILNQNHYDKKSNELIKLSKEFAHDEFOKLFEVLKREAGVKNR 268

Search completed: December 2, 2002, 12:16:00
Job time : 8.97495 secs

THIS PAGE BLANK (USPTO)

lacto-N-neotetraose biosynthesis glycosyl transferase lgta NMB1929 [imported] - Neisseria
C:Species: Neisseria meningitidis
C>Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: D81027
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickey, E.K.; Halt, D.H.; Salberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
Sci, H.; Qiu, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; MUID:2015755; PMID:10710507
A:Accession: D81027
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-349 <TEXT>
A:Cross-references: GB:AE002541; GB:AE002098; NID:g7227175; PIDN:AAF42258.1; PID:g722718
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1929
C:Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 88.9%; Score 1621.5; DB 2; Length 349;
Best Local Similarity 92.5%; Pred. No. 5.2e-127;
Matches 309; Conservative 14; Mismatches 10; Indels 1; Gaps 1;
QY 1 MQLPVSLICAYNVEKFFAOSLAAYVNOTWRNLDILVDDGSTDGTALAKDFQRDSRI 60
17 LQPLVSLICAYNVEKFFAOSLAAYVNOTWRNLDILVDDGSTDGTALAKDFQRDSRI 76
DB 17 LQPLVSLICAYNVEKFFAOSLAAYVNOTWRNLDILVDDGSTDGTALAKDFQRDSRI 76
QY 61 KILQAQNSGLIPSLNIGLELAKSGGGGEYIARTDADDIASPGWIEKIVGEMKDRSI 120
77 RILQAQNSGLIPSLNIGLELAKSGGGGEYIARTDADDIASPGWIEKIVGEMKDRSI 135
DB 77 RILQAQNSGLIPSLNIGLELAKSGGGGEYIARTDADDIASPGWIEKIVGEMKDRSI 135
QY 121 IAMGAMLEVLSEEDGNRLARHHKHKIKMKPTRHEDIAAFPPGPNINHTMTIMRSVY 180
136 IAMGAMLEVLSEEDGNRLARHHKHKIKMKPTRHEDIAAFPPGPNINHTMTIMRSVY 195
DB 136 IAMGAMLEVLSEEDGNRLARHHKHKIKMKPTRHEDIAAFPPGPNINHTMTIMRSVY 195
QY 181 DGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSEKSHVROHEIAOGI 240
196 DGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSEKSHVROHEIAOGI 255
DB 196 DGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSEKSHVROHEIAOGI 255
QY 241 OKTARNDFLOSMGFKTRPDSLEYRQTKAAAYELPEKDLPEDEFERARFLYOCFKRTDTP 300
256 OKTARNDFLOSMGFKTRPDSLEYRQTKAAAYELPEKDLPEDEFERARFLYOCFKRTDTP 315
DB 256 OKTARNDFLOSMGFKTRPDSLEYRQTKAAAYELPEKDLPEDEFERARFLYOCFKRTDTP 315
QY 301 PSGAMLDFAADGRRMRRLFTLRQYFGILRLIKNR 334
316 PSGAMLDFAADGRRMRRLFTLRQYFGILRLIKNR 349
DB 316 PSGAMLDFAADGRRMRRLFTLRQYFGILRLIKNR 349

RESULT 3
S70813
glycosyl transferase A (EC 2.4.-.-) - Neisseria meningitidis
C:Species: Neisseria meningitidis
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S70813
R:Jennings, M.P.; Hood, D.W.; Peak, I.R.A.; Virji, M.; Moxon, E.R.
Mol. Microbiol. 18, 729-740, 1995
A:Title: Molecular analysis of a locus for the biosynthesis and phase-variable expressio
A:Reference number: S70812; MUID:96414473; PMID:8817494
A:Accession: S70813
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <TEXT>
A:Cross-references: EMBL:U25839; NID:g973183; PIDN:AAC4084.1; PID:g973185
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C:Genetics:
A:Gene: lgta
A:Start codon: TTG
C:Superfamily: Neisseria meningitidis glycosyl transferase A
C:Keywords: glycosyltransferase

Query Match 88.7%; Score 1617.5; DB 1; Length 333;

Best Local Similarity 92.5%; Pred. No. 1.1e-126;
Matches 309; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

QY 1 MQLPVSLICAYNVEKFFAOSLAAYVNOTWRNLDILVDDGSTDGTALAKDFQRDSRI 60
17 LQPLVSLICAYNVEKFFAOSLAAYVNOTWRNLDILVDDGSTDGTALAKDFQRDSRI 60
DB 17 LQPLVSLICAYNVEKFFAOSLAAYVNOTWRNLDILVDDGSTDGTALAKDFQRDSRI 60
QY 61 KILQAQNSGLIPSLNIGLELAKSGGGGEYIARTDADDIASPGWIEKIVGEMKDRSI 120
77 RILQAQNSGLIPSLNIGLELAKSGGGGEYIARTDADDIASPGWIEKIVGEMKDRSI 119
DB 77 RILQAQNSGLIPSLNIGLELAKSGGGGEYIARTDADDIASPGWIEKIVGEMKDRSI 119
QY 121 IAMGAMLEVLSEEDGNRLARHHKHKIKMKPTRHEDIAAFPPGPNINHTMTIMRSVY 180
120 IAMGAMLEVLSEEDGNRLARHHKHKIKMKPTRHEDIAAFPPGPNINHTMTIMRSVY 179
DB 120 IAMGAMLEVLSEEDGNRLARHHKHKIKMKPTRHEDIAAFPPGPNINHTMTIMRSVY 179
QY 181 DGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSEKSHVROHEIAOGI 240
180 DGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSEKSHVROHEIAOGI 239
DB 180 DGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSEKSHVROHEIAOGI 239
QY 241 OKTARNDFLOSMGFKTRPDSLEYRQTKAAAYELPEKDLPEDEFERARFLYOCFKRTDTP 300
240 OKTARNDFLOSMGFKTRPDSLEYRQTKAAAYELPEKDLPEDEFERARFLYOCFKRTDTP 299
DB 240 OKTARNDFLOSMGFKTRPDSLEYRQTKAAAYELPEKDLPEDEFERARFLYOCFKRTDTP 299
QY 301 PSGAMLDFAADGRRMRRLFTLRQYFGILRLIKNR 334
300 PSGAMLDFAADGRRMRRLFTLRQYFGILRLIKNR 333
DB 300 PSGAMLDFAADGRRMRRLFTLRQYFGILRLIKNR 333

RESULT 4
H64130
glycosyl transferase homolog H11578 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C>Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C:Accession: H64130
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodek, A.; Kelley, J.M.; Weidman
; D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhmann, J.L.; Geoghegan, N.S.M.
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Venter
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350650; PMID:7542800
A:Accession: H64130
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-323 <TEXT>
A:Cross-references: GB:U32832; GB:IA2023; NID:g1574421; PIDN:AAC23227.1; PID:g1574422
C:Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 33.5%; Score 611.5; DB 1; Length 323;
Best Local Similarity 41.4%; Pred. No. 3.8e-43;
Matches 137; Conservative 62; Mismatches 113; Indels 19; Gaps 5;

QY 3 PLVSVLICAYNVEKFFAOSLAAYVNOTWRNLDILVDDGSTDGTALAKDFQRDSRI 62
5 PLVSVLICAYNVEKFFAOSLAAYVNOTWRNLDILVDDGSTDGTALAKDFQRDSRI 64
DB 5 PLVSVLICAYNVEKFFAOSLAAYVNOTWRNLDILVDDGSTDGTALAKDFQRDSRI 64
QY 63 LQAQNSGLIPSLNIGLELAKSGGGGEYIARTDADDIASPGWIEKIVGEMKDRSI 122
65 ISKNYINIGFINSINIGL-----GCFSGKTFARMADADIAKPSWIEKIVGEMKDRSI 118
DB 65 ISKNYINIGFINSINIGL-----GCFSGKTFARMADADIAKPSWIEKIVGEMKDRSI 118
QY 123 MGAMLEVLSEEDGNRLARHHKHKIKMKPTRHEDIAAFPPGPNINHTMTIMRSVY -D 181
119 MGAMLEVLSEEDGNRLARHHKHKIKMKPTRHEDIAAFPPGPNINHTMTIMRSVY 177
DB 119 MGAMLEVLSEEDGNRLARHHKHKIKMKPTRHEDIAAFPPGPNINHTMTIMRSVY 177
QY 182 GGLRYTERDMADYQFWYDVSKLGRLAYYPALVYRLHANOVSSEKSHVROHEIAOGI 241
178 HGLIFNNDYADYADYKFWSEVSRGLGCLANYPALVYRLHANOVSSEKSHVROHEIAOGI 237
DB 178 HGLIFNNDYADYADYKFWSEVSRGLGCLANYPALVYRLHANOVSSEKSHVROHEIAOGI 237
QY 242 KTAARNDFLOSMGFKTRPDSLEYRQTKAAAYELPEKDLPEDEFERARFLYOCFKRTDTP 301
238 RENITTYLTKIGIDIKYIN-----SVSLLEIYHVDKSNKVLKSLIYEMYSIDKYT 288
DB 238 RENITTYLTKIGIDIKYIN-----SVSLLEIYHVDKSNKVLKSLIYEMYSIDKYT 288
QY 302 SGAMLDFAADGRRMRRLFTLRQYFGILRLIK 332

Db 289 ITSLEHFIK--YHLEFLDKQNLKIKKFFIR 317

RESULT 5

G95948

probable glycosyltransferase protein Smb21189 [imported] - *Sinorhizobium meliloti* (strain C:Species: *Sinorhizobium meliloti*)

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: G95948

R:Finan, T.M.; Weiner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernan Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb *psymb* megaplasmid from the N2-fixing endo A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: G95948

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-367 <KUR> A:Cross-References: GB:AL591985; PIDN:CAC49255.1; PID:g15140741; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid *psymb*

R:Galbert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, hebaute, P.; Vandenbol, M.; Vorholter, F.J.; Weiner, S.; Wells, D.H.; Wong, K.; Yeh, K. A:Title: The composite genome of the legume symbiont *Sinorhizobium meliloti*.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: Smb21189

A:Genome: plasmid

Query Match 17.9%; Score 327; DB 2; Length 367; Best Local Similarity 29.5%; Pred. No. 2e-19;

Matches 103; Conservative 68; Mismatches 128; Indels 50; Gaps 14;

Db

3 PLVSLICAYNWEKYEFAQSIAAVVNOTWRNLDLIYDGSSTDTGLAIKDFOKRDSRIKI 62

4 PVVSYLPVYMEPIYAAIAIESVLRODYELVIAIDGSTDSRDLERKSDSRVSI 63

63 LQAQNSGLIPSLNIGLDLAKSGGGGEYIARTDADIASPGWIEKIVGEMEKRSIIA 122

64 ISR-ENRGDILATINEGL-ALAK-----GELIARMADADIALPERSRLSQAVALFSAEPRAL 116

123 MGWMLVLESEKQGNRLARHHKGIKTKRPTREHDIAPFPENPHNNTMIRRSVI-D 181

117 SGGIDMLI---GNRIIR-GKPNPIY-RPGSRLILSMFEET--FMHSTVYVYNNRVPIE 167

182 GGLRYDTERDMADYOFWYDVSKLGRLAYYPEALVYKRLHANOVSSKH--SVROHEIAOG 239

168 EMARVDPNVYHADFDLFRIRIADRFVHMIDELVAYRIHEDSVTSKHKRQMKRTILTY 227

240 IQKTARNDFLOSMGFKTRPDSLEYRQTKAAAYELP--EKDLEEDDEFERARR----- 288

228 AEMIALDALRLDSALAEIGAAVTSETVARLADLVLAEEISIAOGVEYRRAVEDGALCF 287

289 --FLYGCFKRTDP-----PSGAWLDFADGRRRLFTLRQYTGIL 327

288 EYFLYOLIAEEQPRULTHEFLRTGKW-----GLIRR---RERYGL 326

Db

RESULT 6

B97168

glycosyltransferase [imported] - *Clostridium acetobutylicum*

C:Species: *Clostridium acetobutylicum*

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: B97168

R:Moiling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium *Clostridium acetobutylicum*

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97168

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-333 <KUR>

A:Cross-References: GB:AE001437; PIDN:AAK80133.1; PID:g15025169; GSPDB:GN00168

A:Experimental source: *Clostridium acetobutylicum* ATCC824

C:Genetics:

A:Gene: CAC2175

Query Match 16.1%; Score 293.5; DB 2; Length 333; Best Local Similarity 29.7%; Pred. No. 1e-16;

Matches 80; Conservative 57; Mismatches 97; Indels 35; Gaps 10;

Db

4 LVSVLICAYNWEKYEFAQSIAAVVNOTWRNLDLIYDGSSTDTGLAIKDFOKRDSRIKI 63

1 MSIVMPVYNCERKYLEESTIESLIKOTYRPFELIYWDGSDNSIDINKYANDNRIVYV 60

64 AQAQNSGLIPSLNIGLDLAKSGGGGEYIARTDADIASPGWIEKIVGEMEKRSIIA 123

61 SRDNMGWYVSLNEGIDR-AK-----GSYVARMDADIALPERFERQIEYLNKNKDVIDL 114

124 GAMLEVLSEKQGNRLARHHKGIKTKR--TRHEIAPFPENPHNNTMIRRSVI-D 181

115 ACKVEAFGVDSREKLERH---WYNDLNNSEISIEFLFENCYIAHPSVMMKSYLK 169

A:Accession: B97168

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-333 <KUR>

A:Cross-References: GB:AE001437; PIDN:AAK80133.1; PID:g15025169; GSPDB:GN00168

A:Experimental source: *Clostridium acetobutylicum* ATCC824

C:Genetics:

A:Gene: CAC2175

Query Match 16.1%; Score 293.5; DB 2; Length 333; Best Local Similarity 29.7%; Pred. No. 1e-16;

Matches 80; Conservative 57; Mismatches 97; Indels 35; Gaps 10;

Db

4 LVSVLICAYNWEKYEFAQSIAAVVNOTWRNLDLIYDGSSTDTGLAIKDFOKRDSRIKI 63

1 MSIVMPVYNCERKYLEESTIESLIKOTYRPFELIYWDGSDNSIDINKYANDNRIVYV 60

64 AQAQNSGLIPSLNIGLDLAKSGGGGEYIARTDADIASPGWIEKIVGEMEKRSIIA 123

61 SRDNMGWYVSLNEGIDR-AK-----GSYVARMDADIALPERFERQIEYLNKNKDVIDL 114

124 GAMLEVLSEKQGNRLARHHKGIKTKR--TRHEIAPFPENPHNNTMIRRSVI-D 181

115 ACKVEAFGVDSREKLERH---WYNDLNNSEISIEFLFENCYIAHPSVMMKSYLK 169

182 --GGLRYDTERDMADYOFWYDVSKLGRLAYYPEALVYKRLHANOVSSKH--SVROHEIAOG 238

170 ALGQVYLNKR--TEDYINMLARAIAGKYIAMEEKIMKIRLH-----NDSKITHRAE 220

239 GIQKTARNDFLOSMGFKTRPDSLEYRQTK 267

221 GF--SSIRDIQS-----RLEYVKEK 239

Db

RESULT 7

T44330

glycosyl transferase homolog [imported] - *Vibrio cholerae*

C:Species: *Vibrio cholerae*

C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000

C:Accession: T44330

R:Yamasaki, S.; Shimizu, T.; Hoshino, K.; Ho, S.T.; Shimada, T.; Nait, G.B.; Takeda, Gene 237, 321-332, 1999

A:Title: The genes responsible for O-antigen synthesis of *Vibrio cholerae* O139 are cl

A:Reference number: 222749; MUID:99453293; PMID:10521656

A:Accession: T44330

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-340 <YAN>

A:Cross-References: EMBL:AB012957; NID:g4115688; PIDN:BA433634.1; PID:g3721684

A:Experimental source: strain 022

C:Genetics:

A:Note: wblc

C:Superfamily: Neisseria meningitidis glycosyl transferase A

Query Match 15.3%; Score 279.5; DB 2; Length 340; Best Local Similarity 28.4%; Pred. No. 1.6e-15;

Matches 82; Conservative 55; Mismatches 109; Indels 43; Gaps 9;

Db

3 PLVSLICAYNWEKYEFAQSIAAVVNOTWRNLDLIYDGSSTDTGLAIKDFOKRDSRIKI 62

6 PRIVIMSYVNGEKYLAQAIESTINOTFSDFEITIVDDGSTDSISIIQAYMKDRIYL 65

63 LQAQNSGLIPSLNIGLDLAKSGGGGEYIARTDADIASPGWIEKIVGEMEKRSIIA 122

66 ISRV-NKGLPYSLENAI-SVSKA-----NYIAMMDADDISLPERLETQALVAMENNDICV 118

123 MGWMLVLESEKQGNRLARHHKGIKTKRPTREHDIAPFPENPHNNTMIRRSVI-D 182

119 CGRLAVLFRETSKNNMCH-----PEDHDSIITILLESVCGIHHVVMIRKSVLD- 168

183 GGLRYDTERDM--AEDYQFVYDVSKLGRLAYYPEALVYKRLHANOVSSKH--SVROHEIAOG 240

169 QLDYVYNNFNSODYELMSRIAETKRFYTIQKPLFLFYR-----DTPDGI 213

Db

QY 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEDEPERARRF 289
 | | | | | |
Db 214 TSKVNMND-----GFNKRFPVSVQGTK----QLEKIGLTKNDEMIMHF 253

RESULT 8

AG2188
hypothetical protein alr3062 [imported] - Nostoc sp. (strain PCC 7120)
C:Species: Nostoc sp.
A:Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C:Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C:Accession: AG2188
R:Kaneko, T.; Nakamura, Y.; Molk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriyuchi, N.; Shimpo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
DNA Res. 8, 205-213, 2001
A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium Anabaena PCC 7120
A:Reference number: AB1807; MUID:21595285; PMID:11759840
A:Accession: AG2188
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-3321 <KUR>
A:Cross-references: GB:BA000019; PIDN:BAW74761.1; PID:g17132156; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:
A:Gene: alr3062

Query Match 15.1%; Score 276; DB 2; Length 321;

Best Local Similarity 26.6%; Pred. No. 2.8e-15;

Matches 88; Conservative 63; Mismatches 138; Indels 42; Gaps 12;

QY 3 PLVSVLICAVNVEKYPFAOSLAAVVNOTWRNLDILIVDGSSTDTGLAIADQKRSRIKI 62
 | | | | | | | | | | |
Db 2 PKSVYIPANAPYLPETLESVLRQTHDFEVVYVNDSSDTEWVS--QILDEPRKL 59

QY 63 LAQONSGLIPSLNIGLDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 122
 | | | | | | | | | | |
Db 60 ISOA-MOGLAGARNTGT-----VNASGEYIAFLDADDIWEPYTKLAKOVSVLDENPTVGL 112

QY 123 MGAMLEVLSEK-KDGNRLARHKGKIMKPTRHEDIAAFPPGPNPHNTIMIRSV 181
 | | | | | | | | | | |
Db 113 VYVWVAIVIDQSGSTGKTRKNOVEGYVWPOLTEHNIEC-----GSVALVRRVCFE 163

QY 182 G-GLRYDTERDMAEDYQFWYDVSKLGRLAYYPEALVKYRLHANQVSKHSVRQHEIA 240
 | | | | | | | | | | |
Db 164 KMGLEFRNIGSYVEDMDMLRITATSYDEKVVKEALVYRQRNSAKNNMAMHSAIYI 223

QY 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAAYE--LPEKDLP--EBDEPERARFLYQCFK 295
 | | | | | | | | | | |
Db 224 EK-----AFATASODLQVLKNSYGYTYLCIAMKPLQSFQDKYOKSREF---COQ 270

QY 296 RTDTPPSGAWLDFADG-RMRRLFTLROYFG 325
 | | | | | | | | | | |
Db 271 AVAYYPS---LRSQETIRLSAINIMKRFPG 298

RESULT 9

H97167
glycosyltransferase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: H97167
R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:2159325; PMID:2159325
A:Accession: H97167
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-333 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80131.1; PID:g15025167; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:
A:Gene: CAC2173

Query Match 14.9%; Score 272; DB 2; Length 333;

Best Local Similarity 27.1%; Pred. No. 6.4e-15;

Matches 83; Conservative 72; Mismatches 117; Indels 34; Gaps 12;

QY 5 VSVLICAVNVEKYPFAOSLAAVVNOTWRNLDILIVDGSSTDTGLAIADQKRSRIKI 64
 | | | | | | | | | | |
Db 4 VSVVMMVYNSERKLSIESILNQSISDLFFIINDGSTDNSEKIKKEAKDKRINVIS 63

QY 65 QAONSGLIPSLNIGLDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 124
 | | | | | | | | | | |
Db 64 R-ENKGIYVSLNBAI-RLAK-----GEYIARMDADDISAPKRIEQLSPKSHRIDILG 116

QY 125 AMLEVLSEK-----EEDGNRLARHKGKIMKPTRHEDIAAFPPGPNPHNTIMIRSV 178
 | | | | | | | | | | |
Db 117 IQYKVVGNISNDIKENKEL--NIEFDIY-DNNEKILNTWYCLAH---SVYFRKD 168

QY 179 VIDGRLRYDTERDMAEDYQFWYDVSKLGRLAY-YPEALVKYRLHANQVSKHSVRQHEIA 237
 | | | | | | | | | | |
Db 169 IIRELKGYNDFK--SEDDLMLRAITSGFKIYKRLKELIYFRHESKTRVDNQTGK 226

QY 238 OGIOKTARNDFLOSMGFKTRFDSLEYRQTKAAAYELPEKDLPEDEPERARFLYQCFKRT 297
 | | | | | | | | | | |
Db 227 DGIKIKLIDVFKRE--FKKDFKIYVWASNGGKI--TKVELDFEPKES-----CQIAFV 276

QY 298 DTPPSG 303
 | | | | | | | | | | |
Db 277 DKFKTG 282

RESULT 10

H97168
glycosyltransferase [imported] - Clostridium acetobutylicum
C:Species: Clostridium acetobutylicum
C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C:Accession: A97168
R:Rolling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee, J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium acetobutylicum
A:Reference number: A96900; MUID:2159325; PMID:2159325
A:Accession: A97168
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-336 <KUR>
A:Cross-references: GB:AE001437; PIDN:AAK80132.1; PID:g15025168; GSPDB:GN00168
A:Experimental source: Clostridium acetobutylicum ATCC824
C:Genetics:
A:Gene: CAC2174

Query Match 14.9%; Score 271; DB 2; Length 336;

Best Local Similarity 28.6%; Pred. No. 7.8e-15;

Matches 75; Conservative 64; Mismatches 105; Indels 18; Gaps 10;

QY 3 PLVSVLICAVNVEKYPFAOSLAAVVNOTWRNLDILIVDGSSTDTGLAIADQKRSRIKI 62
 | | | | | | | | | | |
Db 5 PKSVYIPYNSRYLAELIESTIDQTYNDFEITVDSDSTBESYITISYANKQRIIV 64

QY 63 LAQONSGLIPSLNIGLDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 122
 | | | | | | | | | | |
Db 65 ISR-EHRGIVDSINLEGIN-IAR-----GAYIARMDADDISINNRLEKQEFLELNKDVID 117

QY 123 MGAMLEVLSEKDGANLARKHKGKIMKPTRHEDIAAFPPGPNP-HNNTIMIRSV 181
 | | | | | | | | | | |
Db 118 LGRIRIAFDIDIEKQRTI---YNSAFSTKPSQNTIEQVLTSCALPHRSVVEFKDSIYK 173

QY 182 GGLR-VYDTERDMAEDYQFWYDVSKLG-RLAYYPEALVKYRLHANQVSKHSVRQHEIA 239
 | | | | | | | | | | |
Db 174 --LRGTRKEYDIAEDYDLMLRAIRNGYKIVRDECLIKTRVH--NKSIAVEHNPRAVEY 230

QY 240 IOKTARNDFLOSMGFKTRFDSL 261

Db 231 TMK-AKIDYINDTNKKKVDYL 251

RESULT 11

B84114 exopolysaccharide biosynthesis BH3714 [Imported] - *Bacillus halodurans* (strain C-125)

C:Species: *Bacillus halodurans*

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: B84114

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; Hira

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* and

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: B84114

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-271 <STO>

A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA807433.1; GSPDB:GN00

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3714

Query Match 14.2%; Score 258; DB 2; Length 271;

Best Local Similarity 31.5%; Pred. No. 7e-14;

Matches 64; Conservative 45; Mismatches 72; Indels 22; Gaps 7;

1 MOPVSVLCAVNEKFFQSLAAVYVQTRNDLILVDDSGSDGTLAKDFQRKDSRI 60

1 MNPVSVLCAVNEKFFQSLAAVYVQTRNDLILVDDSGSDGTLAKDFQRKDSRI 60

1 MNPVSVLCAVNEKFFQSLAAVYVQTRNDLILVDDSGSDGTLAKDFQRKDSRI 60

61 KILQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADIASPGWIEKIVGEMEKDRSI 120

61 RLTHNNKNGLSYSLAEGV-SLAKA-----PWIAEMDADVDSPKDLAQMOMVKAHSEL 114

121 IAMGAWLEVLSEKDGDNRLARRHHKGIKKPTRHEDIAAF--PFGNPINNTMIMRR 177

115 DILGSY--YIDIDDKGNEL-----EIRKVPTRHKEIANLWTCPE--IHPTVLFKRD 162

178 SVIDGLRYDTERDMAEDYQFWY 200

163 SIHKAG-STDRNLRRQDYDLMF 184

Db 163 SIHKAG-STDRNLRRQDYDLMF 184

RESULT 12

AI2091

glucosyltransferase [Imported] - *Nostoc* sp. (strain PCC 7120)

C:Species: *Nostoc* sp.

A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AI2091

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Iriguchi

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S

DNA Res. 8, 205-213, 2001

A:Title: Complete genomic sequence of the filamentous Nitrogen-fixing Cyanobacterium *Ana*

A:Reference number: AB1807; MUID:21595285; PMID:11759840

A:Accession: AI2091

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-343 <KUR>

A:Cross-references: GB:BA000019; PIDN:BA873987.1; PID:g17131380; GSPDB:GN00179

A:Experimental source: strain PCC 7120

C:Genetics:

A:Gene: all2288

Query Match 13.9%; Score 252.5; DB 2; Length 343;

Best Local Similarity 28.8%; Pred. No. 2.8e-13;

Matches 84; Conservative 63; Mismatches 100; Indels 45; Gaps 13;

5 VSVLCAVNEKFFQSLAAVYVQTRNDLILVDDSGSDGTLAKDFQRKDSRIKILA 64

3 LSVLIVYNESSVAETLRSVLAQTYRDLIELIVDDSGSDKSIDICKOFO--DERIRIHH 60

Qy 65 QAQNSGLIPSLNIGDELAKSGGGGEYIARTDADIASPGWIEKIVGEMEKDRSIANG 124

61 Q-QNRLAGARNTGIRQ-----ARCEYLAFVSDDLMLPEKLAKLHEHREPEVGEF 113

Qy 125 ANLEVLSEKDGDNRLARRHHKGIKKPTRHEDIAAFPP-CPPIIN-NTMIRBSV 179

114 SRSLIDQ-----GKPLGIYQMP-KLTPITPEYLCRNPDIANGSSVYIRAV 160

Qy 180 ID-----GGRLRYDTERDMAEDYQFWYDV-S-KLGRLAYPEALVYRLHANGV 225

161 LDTIKQENLVGEVEDFYFDSDRFQSEDIQWLRILAQTTWKIEGIPALTYLRVMGGL 220

Qy 226 SSKHSYRQHEIAQGI--OKTARN-DFLOSGMGFTRPDSLEYROTAKAAYELP 274

221 SA-NVLKQYESMRILVKNQAVNPEFIERMGNNAKAYQRLYARRTRRSP 271

Db 221 SA-NVLKQYESMRILVKNQAVNPEFIERMGNNAKAYQRLYARRTRRSP 271

RESULT 13

AB4107

glycosyltransferase BH3657 [Imported] - *Bacillus halodurans* (strain C-125)

C:Species: *Bacillus halodurans*

C>Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001

C:Accession: AB4107

R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H

Nucleic Acids Res. 28, 4317-4331, 2000

A:Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a

A:Reference number: A83650; MUID:20512582; PMID:11058132

A:Accession: AB4107

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-732 <STO>

A:Cross-references: GB:AP001519; GB:BA000004; NID:g10176109; PIDN:BA807376.1; GSPDB:G

A:Experimental source: strain C-125

C:Genetics:

A:Gene: BH3657

Query Match 13.5%; Score 247; DB 2; Length 732;

Best Local Similarity 24.9%; Pred. No. 2.3e-12;

Matches 87; Conservative 59; Mismatches 114; Indels 90; Gaps 12;

1 MOPVSVLCAVNEKFFQSLAAVYVQTRNDLILVDDSGSDGTLAKDFQRKDSRI 60

109 MEKVTVAIIPAYNSEDEGIATSIESIQQTWTNIEILVDDCSYDGTALIKRYAASDDRY 168

Qy 61 KILQAQNSGLIPSLNIGDELAKSGGGGEYIARTDADIASPGWIEKIVGEMEKDRSI 120

169 KYLKEINSGAVYARNIAQQAA-----GEVTLINDADWSHPAKIEYQVNHLEHPY 222

Qy 121 IAMGAWLEVLSEKDGDNRLARRHHKGIKKPTRHEDIAAFPPGKNP-----IHNTMIM 175

223 IANTSOQAKMTEB-----LTFPRKGRGEXIYVNMSSLMF 257

Qy 176 RRSVIDGLRYDTERDMAEDYQFWYVSKL-GRLAYYPEALVYK--RLHANOVSSKHSV 231

258 RKNPVAMENIGFMFVSVFAGADGEFRKRLKVFGEKAVAVDASGVSFQROTGSILTGNAHF 317

Qy 232 RQHEINQIOKTAARNDFLOSMGFTRPDSLE--YROTAAAYVLPKQD-----LPEDDER 285

318 GFHGYFMGVR-----KEYFDSISYHRYAESLRDPEPRFLFAVPEPMPO 364

Qy 286 ARREFLYQCFKRTDTPPSGAWLDFADGRMRRLFTLQYFGIL---YRLI 331

365 RE-----KKT-----ADG-----RRHFDVILASFRLL 387

Db 365 RE-----KKT-----ADG-----RRHFDVILASFRLL 387

RESULT 14

AH2189

hypothetical protein alr3071 [Imported] - *Nostoc* sp. (strain PCC 7120)

C:Species: *Nostoc* sp.

A:Note: *Nostoc* sp. strain PCC 7120 is a synonym of *Anabaena* sp. strain PCC 7120

C>Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002

C:Accession: AH2189

R:Kaneko, T.; Nakamura, Y.; Wolk, C.P.; Kuritz, T.; Sasamoto, S.; Watanabe, A.; Irigun

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S.
 DNA Res. 8, 205-213, 2001
 A:Title: Complete Genomic Sequence of the Filamentous Nitrogen-fixing Cyanobacterium *Anabaena*
 A:Reference number: AB1807; MUID:21595285; PMID:11759840
 A:Accession: AH2189
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-318 <KUR>
 A:Cross-references: GB:BA000019; PIDs:BAB74770.1; PIDs:g17132165; GSPDB:GN00179
 A:Experimental source: strain FCC 7120
 C:Genetics:
 C:Gene: alr3071

Query Match	13.5%;	Score 245.5;	DB 2;	Length 318;
Best Local Similarity	28.9%;	Pred. No. 9.5e-13;		
Matches 70;	Conservative 45;	Mismatches 108;	Indels 19;	Gaps 5;

Qy 3 PLYSVLLCAYNNEKYYFAOSLAAVYNGTWRNDLILIVDGGSTDEGLAKDKFQRDRSRIKI 62
| | | | | | | | : : : : : | | | | | | | : : : : :
Db 2 PKYSVLIPATYNAKKYLPAIVEASVLAQOSTDIETLIILINGSSDN--IIAMTAQTIDPRVOV 59

```

0Y      LAQANSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIA 122
      ::| | | | : | : | | | | | : | | | : :
63      ISQ-QNQGSLGARNTGIHH-----ASGEYIAFIADADDMLPTKLEKQVCLDNSPQAG 112
      60
Db

```

```

Oy      123  MGAMLEVLSEE-KDCGRRLARHHKGGKRWKKPTRHEDIAAFPEPCGNPIHNNTMIMRSVID 181
          :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db      113  VYTWTAWTDGKPTGCVIVASHVEGYWQMVVNDKIS-----NGSSAMVRICFD 163

```

```
Qy      182 GGLRYDTERMAEDYQFWYDVSKLGRLAYYPBALVKYRLHANOVSSKHSVQHEINOIGIO 241
        :||| :|| |:: | | | | : :| | :
Db      164 KVGLEDTELTSSDRMMIRLAHYHFAVVKEPITLYRRHSQSMKNRPKMLKNIRVFEE 223
```

QY	242	KT	243
Db	224	KT	225

RESULT 15
AE2160
glycosyltransferase alr2836 [imported] - Nostoc sp. (strain PCC 7120)

A: Note: Nostoc sp. strain PCC 7120 is a synonym of Anabaena sp. strain PCC 7120
C: Date: 14-Dec-2001 #sequence_revision 14-Dec-2001 #text_change 30-Jun-2002
C: Accession: AE2160

Nakazaki, N.; Shimo, S.; Sugimoto, M.; Takazawa, M.; Yamada, M.; Yasuda, M.; Tabata, S. DNA Res. 8: 205-213, 2001

A:Accession: AE2160
A:Status: preliminary
A:Molecule type: DNA

A:Cross-references: GB:BA000019; PIDN:BAW74535.1; PID:g17131930; GSPDB:GN00179
A:Experimental source: strain PCC 7120
C:Genetics:

Query Match	13.3%	Score 243;	DB 2;	Length 322;
Best Local Similarity	28.5%	Pred. No. 1.66-12;		
Matches 96; Conservative	61;	Mismatches 112;	Indels 68;	Gaps 19;

```
QY      5  VSVLICAYNVEKYFAQSLAAVNNQTRNMDILIVDDGSTDGTALIAKFGQR-DSRIKIL 63
      ::::| | :| :::::| | :| ::::| | | | | | | | :| :::| | :
Db      3  ISVITSNNYARYLSRAINSVLAQTHSDIEIVIVDDGSTDNSRDVITLOEQAPDKIKPI 62
```

Oy 64 AQAQNSGLIPSLNIGHDELAASGGGGEYIARTDADDIASPGWTEKIVGEMEKDRSIIM 123
 ||| : | | ||| : | ::|| : : :
Db 63 FGA-NQGCGAFAFNAG-----AAATEVVALFDADDDVKPHKILORIV-EVFQTSDVGV 114

0y 124 GAMLEVLSEKDG--RLARHHKGLWKPPRHEDIAF-----PPEGNP1HNNT 172
 ::: ||| : : | ||: | :

```

Db 115 MHHDIT---DGMDKITDASTG-----PKLSEDLASVILDTGNMCP-----PTSG 160
Qy 173 MIMRSYIDGCLARYDTE--RDMADYOFWYDVSXKIGRLAYVPEALVYRYRLHA--NOVSSK 228
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 161 LAYREVLEKEFPIDPYKWRITMA-DGCLITCTATLGNKIKTLOELAYLRIHGANNHMSAA 219
Qy 229 HSVROHETIAQ--GIQOKTAR--NDFLOSMGEKTRPD---SLEYROTKAAYELPEKDLPE- 260
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 220 SATSBEQEAKSQAGIEIMTNOYINDELVLRIGYARVDLSRNLTQYRRTK--YYQSRQMDUREV 277
Qy 281 -----EDFERA---RRFLYQCFR---RTD 298
   :  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |  |
Db 278 WGISRLLIGWPFYSGQGERAYYLAFELFRSGKFFLLRD 314

```

```
Search completed: December 2, 2002, 12:00:07
Job time : 17.4445 secs
```

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 8.46958 Seconds

(without alignments)
1704.189 Million cell updates/sec

Title: US-10-007-267-3

Perfect score: 1823

Sequence: 1 MQLPVSLICAYNVEKYFAQ.....RLIKRRQARSASAKGEQEI 348

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	611.5	33.5	323	1 YF78_HAEIN	Q57287 haemophilus
2	243	13.3	322	1 YS86_ANASP	P22639 methanobacillus
3	229	12.6	290	1 YA57_METUA	Q58457 methanococcus
4	204.5	11.2	250	1 Y868_HAEIN	Q57022 haemophilus
5	189.5	10.4	344	1 Y1BD_ECOLI	P11290 escherichia
6	187.5	10.3	266	1 AMSE_ERMAN	Q46635 erwinia
7	164	9.0	301	1 AMSB_ERMAN	Q46632 erwinia amy
8	163.5	9.0	446	1 GGAB_BACSU	P46917 bacillus su
9	161	8.8	900	1 GGAB_BACSU	P46918 bacillus su
10	146	8.0	267	1 YG95_HAEIN	Q48215 haemophilus
11	146	8.0	348	1 EXOQ_RHIME	P33697 rhizobium m
12	144.5	7.9	268	1 YNPF_BACSU	P39614 bacillus su
13	143	7.8	256	1 SP8A_BACSU	P39621 bacillus su
14	135	7.4	441	1 YCQ0_ECOLI	P75905 escherichia
15	132.5	7.3	299	1 Y060_MYCPN	P75042 mycoplasma
16	132.5	7.3	909	1 Y4G1_RHISM	P55465 rhizobium s
17	124.5	6.8	341	1 Y33B_MYCPN	P75302 mycoplasma
18	122	6.7	342	1 EXOQ_RHIME	P33700 rhizobium m
19	121.5	6.7	346	1 YP20_MYCTU	Q50587 mycobacteri
20	120.5	6.6	1275	1 RFBG_MTXXA	Q50864 myxococcus
21	120	6.6	299	1 Y025_MYCPN	P75086 mycoplasma
22	118.5	6.5	426	1 CGED_BACSU	P42092 bacillus su
23	117.5	6.4	330	1 EXOA_RHIME	P33691 rhizobium m
24	116.5	6.4	424	1 NODC_RHILIO	P17862 rhizobium l
25	116.5	6.4	559	1 PACT_MOUSE	Q08812 mus musculu
26	114	6.3	297	1 Y060_MYCGE	P47306 mycoplasma
27	114	6.3	319	1 YF18_MYCTU	Q05090 mycobacteri
28	113.5	6.2	559	1 PACT_BOVIN	Q07537 bos taurus
29	113.5	6.2	559	1 PACT_HUMAN	Q10472 homo sapien
30	113.5	6.2	559	1 PACT_RAT	Q10473 rattus norv
31	112.5	6.2	256	1 YF57_MYCTU	Q50459 mycobacteri
32	111.5	6.1	323	1 YKCC_BACSU	Q04319 bacillus su
33	111.5	6.1	426	1 NODC_RHIME	P34341 rhizobium m

34	111.5	6.1	612	1 PAG3_CAEBL	P34678 caenorhabdi
35	110.5	6.1	298	1 Y025_MYCGE	P47271 mycoplasma
36	110.5	6.1	559	1 PACT_PIG	Q29121 sus scrofa
37	109	6.0	289	1 RFBG_SHIFL	P37783 shigella fl
38	108	5.9	329	1 CSBB_BACSU	Q45539 bacillus fl
39	105.5	5.8	241	1 DPM1_DROME	Q9Y1U7 drosophila
40	104.5	5.7	513	1 YTH1_RHOER	P46370 rhodococcus
41	103.5	5.7	236	1 DPM1_SCHPO	Q14466 schizosacch
42	103.5	5.7	279	1 WCAA_ECOLI	P77414 escherichia
43	103	5.7	182	1 NODC_BRABL	P53417 bradyrhizob
44	98.5	5.4	260	1 DPM1_HUMAN	Q60762 homo sapien
45	98	5.4	294	1 YG96_HAEIN	Q48214 haemophilus

ALIGNMENTS

```

RESULT 1
ID YF78_HAEIN          STANDARD;          PRT;          323 AA.
AC Q57287; 005077;
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DE Putative glycosyl transferase H1578 (EC 2.-.-.-).
GN H1578.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=9350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shiley R., Liu L.-T., Glodok A., Kelley J.M.,
RA Weisman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Usterback T.R., Hanna M.C., Nguyen D.T., Saudke D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghagen N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
  Rd.";
RL Science 269:496-512(1995).
CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U3832; AAC23227.1; -.
DR TIGR: H1578; -.
DR InterPro: IPR001173; Glycosyltransferase_2.
DR Pfam: PF00535; Glycosyltransferase;
KW Hypothetical protein; transferase; glycosyltransferase;
KW Complete proteome.
SQ
SEQUENCE 323 AA; 37680 MW; 7C6C2681039A5B4 CRC64;

Query Match          33.5%; Score 611.5; DB 1; Length 323;
Best Local Similarity 41.4%; Pred. No. 6.4e-43;
Matches 137; Conservative 62; Mismatches 113; Indels 19; Gaps 5;

OY 3 PLVSVLICAYNVEKYFAQSLAAVAVNQTWRNLDILYDDSGTGTIAADFGKPSRIKI 62
   |||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||
Db 5 PLVSVVICAYNVEKYFAQSLAAVAVNQTWRNLDILYDDSGTGTIAADFGKPSRIKI 64
   |||||:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||:|:|:|:|||||

OY 63 LQAQNSGLIPSLINIGLDELAKSGGGGGEYIARTADDIASPGWIEKIYGVEMEKDRSIITA 122

```



```
Db 65 ISKKNYLMGFNSLNIGL-----GCFSGKFFAMADADDLAKPSMIKRYWLEKNDHITR 118
123 MGAMLEVLSEKGNRLARHHKHGKIMKRTREDIAAFPPGNPIHNNMIRRSYI-D 181
119 MGSYLEIIEKEG-1IGSYKGTGIDMKPNLNDICEAMLEFYNPHNNMIRKANYRE 177
182 GGLRYDTERDMADYQFWYDVSKLGLRAYYPALVRYRLHANOVSCKSHVROHEIAQIO 241
178 HKLIFNMKDYADYAEKFWSEVSRGLGLANPEALVRYRLHNGNOTSSVYHNEQETAKKIK 237
Db 242 KTRANDPLOSMTKPTFDSLEIYRQTKAAAYELPEKDLPEEDFERARFLYQCFKRTDTP 301
238 RENTTYLNNKIGIDIKVIN-----SVSLIEIYHVDKSNKVLKSLIEMYSIDKXT 288
OY 302 SGAMLEPDAADGRMRRLFTLRYGILRLIK 332
Db 289 ITSLHPIK--YHLEPDLKONKIITKPIR 317
```

RESULT 2

```
YS86 ANASP
ID YS86 ANASP STANDARD: PRT: 322 AA.
AC P22639;
DT 01-AUG-1991 (rel. 19, Created)
DT 15-JUN-2002 (rel. 41, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Putative glycosyl transferase alr2836 (EC 2.-.-.-).
GN ALR2836
OS Anabaena sp. (strain PCC 7120).
OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc.
OX NCBI_Taxid=103690;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21595285; PubMed=11759840;
RA Kaneko T., Nakamura Y., Wolk C.P., Kuritz T., Sasamoto S.,
RA Watanabe A., Iriuguchi M., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kohara M., Matsumoto M., Matsuno A., Muraki A.,
RA Nakazaki N., Shimpo S., Sugimoto M., Takazawa M., Yamada M.,
RA Yasuda M., Tabata S.;
RA "Complete genomic sequence of the filamentous nitrogen-fixing
RT cyanobacterium Anabaena sp. strain PCC 7120.";
RL DNA Res. 8:205-213(2001).
RN [2]
RP SEQUENCE OF 1-131 FROM N.A.
RX MEDLINE=90264305; PubMed=2111805;
RA Holland D., Wolk C.P.;
RA "Identification and characterization of heta, a gene that acts early
RT in the process of morphological differentiation of heterocysts.";
RL J. Bacteriol. 172:3131-3137(1990).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: AF003591; BAB74535.1; -.
DR EMBL: AF031959; AAC32401.1; ALT_INIT.
DR PIR: B35391; B35391.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2.1.
KW Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 322 AA; 36388 MW; 753C2FB59327D968 CRC64;
```

```
Query Match 13.3%; Score 243; DB 1; Length 322;
Best Local Similarity 28.5%; Pred. No. 1e-12;
Matches 96; Conservative 61; Mismatches 112; Indels 68; Gaps 19;
```

```
OY 5 VSVLCAYNVEKRYFAOSLAAYVNTWRNLDILVDGSDTGLTAKDFQR-DSRIKIL 63
Db 3 ISVLIYNVYARLSRAINSVLAQTHSDIEIYVDGSDTNSRDYITQEQAPDKIKPI 62
OY 64 AQONSGILPISNTIGDELAKSGGGGEYIARDADIASPGMIKIEIVEMEKDSIIM 123
Db 63 FQA-NOGOGGAENAF-----AAATGEVVALADADVAKPHKLQIV-EVQTSDVAV 114
OY 124 GAVLEVLSEKGN--RLARHHKHGKIMKRTREDIAF-----FPPGNPIHNN 172
Db 115 MHHLDI-----DNDKTDIOASTQ-----PKLSEDLASVILQTNAMCFP-----PTSG 160
OY 173 MIMRSVIDGLRYDTE--RDMAEDYQFWYDVSKLGLRAYYPALVRYRLHA--NOVSSK 228
Db 161 LAYRREVLEKVPIDPVKKRIMA-DCGIIYCAFLGKIKITLQENLAYYRHGANNHMSA 219
OY 229 HSYRQHEINQ--GIQTKAR--NDPLOSMTKPTFDSLEIYRQTKAAAYELPEKDLPE 280
Db 220 SATSEQAKSQAGIEINTNOYINDELVRIGYGARVDLSRNIQYRRK--YYQSRQMDLREV 277
OY 281 -----EDPERA---RRFLYQCFK---FTD 298
Db 278 WGISRLILGWPFYSGGERAYIARFLFKSGKFLRSD 314
```

RESULT 3

```
YA57 METJA
ID YA57 METJA STANDARD: PRT: 290 AA.
AC 058457;
DT 15-JUL-1998 (rel. 36, Created)
DT 30-MAY-2000 (rel. 39, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Putative glycosyl transferase MJ1057 (EC 2.-.-.-).
GN MJ1057.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_Taxid=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2861 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerelevage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reisch C.I.,
RA Overbeek R., Kirkness E.F., Weissflock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhrmann J.L., Nguyen D.,
RA Uutterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii.";
RL Science 273:1058-1073(1996).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
DR EMBL: U67549; AAB99061.1; -.
DR HSSP: P39621; IQGQ.
DR TIGR: MJ1057; IQGQ.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2.1.
KW Hypothetical protein; Transferase; Complete proteome.
SQ SEQUENCE 290 AA; 35099 MW; 3f6A1B221C420D74 CRC64;
```

```
Query Match 12.6%; Score 229; DB 1; Length 290;
Best Local Similarity 29.9%; Pred. No. 1.3e-11;
```


	Matches	70;	Conservative	50;	Mismatches	72;	Indels	42;	Gaps	10;	
QY	2	OP	PLSVLICAVN-VKRYPAQSLAAV	NOTWRMLD-ILIVD	GS	TG	TLAIAK	QFKR	DSR	59	
		:	::		:		:	::	:	::	
Db	7	KPLVSV	VAVATYEPKRYLKE	SIESIX	NOTXKQF	FXILVD	NPNNKKA	AEIIE	LEYO	OKDKR 66	
QY	60	IKLIAQA	NSGLIPLNTIGLDELAK	SGGGG	GXYIARTAD	IASP	GMIEIYK	GE	MEK	DRS 119	
		:	::		:	::		:		::	
Db	67	IIFINERN	LFGASRNKAVN-IAR	----	GXYII	ILDAD	DIALKRL	LEK	QFY	WENNND 120	
QY	120	IIAMGAM	EVLSSEKDG	NRLARH	RKGI	IKW-K	EPTRH-ED	IAA	FP	PGNPIHNNTM 175	
		:	::		:		:	::	:	::	
Db	121	IDLFS	WYFIDE	-----	NGN	LKFK	PEK	KYFK	EIK	KYF-FKEHLV	HPSPM 167
QY	176	RRSV	LDGGI	RYTER	EDMA	DYQF	W----	YVSK	GR	LAYP	PEALVYKRL 220
		:		::		:		:		:	::
Db	168	VASK	ILTKL	KVDE	KILRS	ODYF	WIR	CJ	ANDY	KFPIIE-EEFL	KRI 213

```

RESULT 4
Y868_HAEIN STANDARD: PRT; 250 AA.
ID Y868_HAEIN STANDARD: PRT; 250 AA.
AC Q57022; P96336;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase H10868 (EC 2.-.-.-).
GN H10868.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxId=727;
RN [1]
RP SEQUENCE FROM N.A.
RP STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merriick J.M.,
RA McEneney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodak A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Overbeek T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Flier I.D., Fitcham J.L., Furman J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.G.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RI Science 269:496-512(1995).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U32768; AAC2526.1; -.
DR HSSP: P39621; 1OGQ.
DR TIGR: H10868; -.
DR InterPro: IPRO01173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 250 AA; 28915 MW; ASD8220129782EE98 CRC64;

Query Match 11.2%; Score 204.5; DB 1; Length 250;
Best Local Similarity 25.9%; Pred. No. 1.le-09;
Matches 63; Conservative 52; Mismatches 93; Indels 35; Gaps 8

3 PLVSLVCATNWKFAQSLAAVNQNTWNLIDLVDGSDGTDLAIADFGKRDSPRIKI 62
||||| | | | : |||||:||||| : | : | | | :

```

```

Db      4 PLSTIMPYNAECTLNGIISLCNASTQNIELLIDDGSTDKSEIILNIIIDKDRVKL 63
QY      63 IAAQNSGLIPSLNTGLDELAKSGGGGGEYIARTDADDIASPGMIKIVGEMEKDRSIIA 122
Db      64 FFTPTNOSPAAARNIGLEK-----AOGDYITFLPDSDFIANDLKEQLNFMLONHLYMT 117
QY      123 WCAMLEVEISEKDGRIARLHHKHGKIKMKPRRHEDIAFFPFGPINNMIMRSYIDG 182
Db      118 HENVAFCC-:::|::|-----DYTLTLD-GNQFKIMTVLYRESIK- 164
QY      183 GLRYTERDMAEDYQFWVDVSKLGRLAYPEALYKRYLHANQVSS-----KHSYRQEIHA 237
Db      165 ILRFPNIKH-EDYAFFLDCK-----EVNQSLIYHQASFPRIKGVSVSSNKKF 211
QY      238 QGI 240
Db      214 SAI 216

```

```

RESULT 5
YIBD_ECOLI ID YIBD_ECOLI STANDARD: PRT; 344 AA.
AC P11290; P2299g;
DT 01-JUL-1989 (Rel. 11, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE putative glycosyl transferase yibD (EC 2.-.-.-).
GN yibD OR B3615.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_Taxid=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE=94316500; PubMed=8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
RN [2]
RP SEQUENCE OF 1-198 FROM N.A.
RC STRAIN-K12;
RX MEDLINE=89174812; PubMed=2647748;
RA Aronson B.D., Somerville R.L., Eperley B.R., Dekker E.E.;
RT "The primary structure of Escherichia coli L-threonine
RT dehydrogenase."
RL J. Biol. Chem. 264:5226-5232(1989).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U00039; AAB18592.1; -
DR EMBL; AE000439; AAC76639.1; -
DR EMBL; X06690; CAA29885.1; -.
DR PIR; A30268; Q3ECTH.
DR PIR; B33276; B33276.
DR Ecocore; EGI1266; yibD.
DR InterPro; IPR001173; Glycosyltransferase_2.
DR Pfam; PF00535; Glycosyltransferase_2_1.
KW Hypothetical protein; Transferase; glycosyltransferase;
KW Complete Proteome.
FT CONFLICT 1 26
FT (IN REF. 2)
FT MUNSINKSVITIPLYNAGDPRFCME -> MRAMISALYWK
FT FT
SQ SEQUENCE 344 AA: 40524 MW; 9B5DAFE86D547A CRC64;

```

Query Match	10.4%;	Score 189.5;	DB 1;	Length 344;
Best Local Similarity	24.2%;	Pred. NO. 2.7e-08;		
Matches	79;	Conservative	55;	Mismatches 132;
			Indels	61;
			Gaps	14;

[illegible]

RESULT

ID	AMSE_ERNAME	STANDARD	PRT	266 AA.
AC	046635:			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Amylovoran biosynthesis glycosyl transferase amse (EC 2.-.-.-).			
GN	AMSE.			
OS	Erwinia amylovora.			
OC	Bacteria.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OX	NCBI_TaxID=552;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=EAI/79;			
RX	MEDLINE=95319333; PubMed=7596293;			
RA	Bugert P., Geider K.;			
RT	"Molecular analysis of the amse operon required for exopolysaccharide synthesis of Erwinia amylovora.";			
CC	Mol. Microbiol. 15:917-933(1995).			
CC	-1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF AMYLOVORAN WHICH FUNCTIONS AS A VIRULENCE FACTOR.			
CC	-1- PATHWAY: Exopolysaccharide biosynthesis.			
CC	-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.			
CC	-----			
CC	THIS SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See http://www.isb-sib.ch/announce/sib.ch).			
CC	or send an email to license@sib-sib.ch).			

DR EMBL; X77921; CAA54886.1; -.
DR InterPro; IPR001173; Glycos_transf_2.

KM Exopolysaccharide synthesis; Transferase; Glycosyltransferase
SQ SEQUENCE 266 AA; 30748 MW; 163268A4210EB47B CRC64;

Query Match	10.3%;	Score 187.5;	DB 1;	Length 266;
Best Local Similarity	25.4%;	Pred. No. 2.8e-08;		
Matches 64;	Conservative 42;	Mismatches 115;	Indels 31;	Gaps 7;

QY	4	LVSVLIAAYNNEK--YFAQSLAAVNOTWNRILILYVDSSTGTLAIAADFQKPSRIK	61
Db	1	MSVSLISLYNEKEPBNLEQCSLESHOOTLADNELVLYVDPVSESLKAAYTRANLLPV	60
QY	62	TLAQAQNSGLIPSLNTIGDELAKSGGGGGEYIARTDADIASPGWTEKIVGEMEKORSII	121
Db	61	IYPLEKNTKGLKALNAGLERCTHN-----VVAAMDPPDLCIPERFEKQISYMSHEPV	114
QY	122	AMGAWLEVLSEKKGNSRLARHNKHK--IYKKPTRHEDIAAFEPFGSPITNNIMBRSV	179
Db	115	LSGA-----AYIEPDEGKERLKLPLSNNDHEFAAMKRPENHMCVFRKDK	162
QY	180	IDDGLRYDTERDMAEDYQFWYDVSKLR-LAYYPEALVYRLHANAQVS-----KHSV	231
Db	163	VYSAGSYQ-HHLUMEDYULYRLIMSLGHPVANLPDVLMTKVAGSDVMNKRGRWNYIKSEV	221
QY	232	QHHEIAQSIQKT 243	
Db	222	QVYRLKLAKQT 233	

RESULT 7

ID	AMSB. ERAWM	STANDARD:	PRT:	301 AA.
AC	046632;			
DT	01-NOV-1997 (Rel. 35, Created)			
DT	01-NOV-1997 (Rel. 35, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Amylovoran biosynthesis glycosyl transferase amsb (EC 2.-.-.-).			
GN	AMSB.			
OS	Erywinia amylovora.			
OC	Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;			
OC	Erywinia.			
OX	NCBI_TaxID=552;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN-EAI/79;			
RX	MEDLINE=9531933; PubMed=7596293;			
RA	Bugert P., Geider K.;			
RT	"Molecular analysis of the amv operon required for exopolysaccharide			
RL	synthesis of Erywinia amylovora.";			
RL	Mol. Microbiol. 15:917-933(1995)."			
CC	-1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF AMYLOVORAN WHICH			
CC	FUNCTIONS AS A VIRULENCE FACTOR. MAY FUNCTION AS A GLYCOSYL			
CC	TRANSFERASE WHICH TRANSFERS GALACTOSE FROM UDP-GALACTOSE TO A			
CC	LIPID-LINKED AMYLOVORAN-SUBUNIT PRECURSOR.			
CC	-1- PATHWAY: Exopolysaccharide biosynthesis.			
CC	-1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@isb-sib.ch).			
CC	-----			
DR	EMBL; X77921; CAA54883.1; -			
DR	InterPro: IPR001173; Glycos_transf_2.			
DR	Pfam: PF00535; Glycos_transf_2: 1.			
KW	Exopolysaccharide synthesis; transferase; Glycosyltransferase.			
SO	SEQUENCE 301 AA; 34788 MW; 99DD0EE3080EECD0 CRC64;			

Query Match	9.08;	Score 10
Best Local Similarity	21.88;	Pred. Nd

6 SYLICAYNEKFEAOSLA VVNQTRNDILTVDDGSTDTLAI AKDFQ

Query Match	9.0%;	Score 164;	DB 1;	Length 301;
Best Local Similarity	21.8%;	Pred. No. 2.8e-06;		
Matches 76;	Conservative 57;	Mismatches 106;	Indels 110;	Gaps 16;

[illegible]

Db 56 ERYGGL--KINLVLSKTLNGAGARRNGICDLATGDYVCFDADD---EWHKDLQNMIS 109
 QY 107 -IERKIVGEMEK-----DRSIIANGAMLEVLSEEDKGRRLARHHKHKIKMKPTFHEDIA 159
 Db 110 LIEKLEGGGDRFRFIITYSVNIIDQSSFLKMPKPKVG-----EHESIA 152
 QY 160 AAFPFG--NPJHNTMTMIRRSVIDG--LRYDTERDMADYQFWYDVSFKLGRLAYVPEALV 216
 Db 153 EYL-EGCYGFIQTSTYLVKRE--DAAEIRFDEYRIHQDDYLCIRADLKG-----F 200
 QY 217 KYRLHANOVSXKSHYRQHEIAGIOKTRANDFLOSMGKRTFDSLEFYRQTAAYELPEK 276
 Db 201 KFMIAIDPLANYHMV-----TRFGS--QHKSGSVKSYLFWL 234
 QY 277 DLPEDEFARARFLYOCFKRTDTPPSGAMLPADGR-MRLSLFTLROYE 324
 Db 235 DAMKPHLTRDYVITYYKAKLP-----LKKMDKSLQSLSPARYF 275

RESULT 8
 GGAB_BACSU STANDARD; PRT; 446 AA.
 ID GGAB_BACSU
 AC P46917;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Minor teichoic acids biosynthesis protein ggaA.
 GN GGA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Borries R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Connerion I.F., Cummings N.J., Daniel R.A.,
 RA Delizot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Ertlan K.D., Errington J., Fabret C., Ferrari E., Follmer D.,
 RA Fritz C., Fujita M., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Gilm S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Halberst H., Holsappel S., Hosono S., Hulio M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapius A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogiwara A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Portwoll S., Prescott A.M.,
 RA Prescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadle Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serro P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Tamakoshi A., Tanaka T., Terstra P., Topponi A.,
 RA Tosato V., Uchiyama S., Vandenberg M., Vannier F., Vassartot A.,
 RA Viari A., Wandut R., Wedler E., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasumoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium Bacillus
 RT subtilis."
 RL Nature 390:249-256(1997).
 CC -I- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF GALACTOSAMINE-CONTAINING
 CC MINOR TEICHOIC ACIDS.

CC -I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC EMBL; U13979; AAA73512.1; -;
 DR EMBL; Z99122; CAB1586.1; -;
 DR Subtilist; BG1367; ggaA.
 DR InterPro; IPR001173; Glycos.transf_2.
 DR Pfam; PF00535; Glycos.transf_2; 1.
 DR Transferase; Glycosyltransferase; Complete proteome.
 KM SEQUENCE 446 AA; 53148 MW; BB3698D96856C42 CRC64;
 SQ

Query Match 9.0%; Score 163.5; DB 1; Length 446;
 Best Local Similarity 19.8%; Pred. No. 5.2e-06;
 Matches 74; Conservative 72; Mismatches 139; Indels 89; Gaps 15;

QY 4 LVSVLICAYNWEKRYFAQSLAAVYVQT---WRNLDILIVDGSSTDTLAIARDFOK-RDSR 59
 Db 1 MFSITIPYNSGNLRYASIESVLTNOSIGFKEINIELILIDGSDVSDSPQICESFKMLYNN 60
 QY 60 IKILAQANSGILPSNLGLDELAKSGGGGEYIARTADDIASGWIETKIVGEMKRS 119
 Db 61 IKIM-KIENSGPSAARNGLSNVSR---SKFIFLSDDAFSONALQSYV-DEFCSE 114
 QY 120 IYAMGAMLEVLSEEDKGRRLARHHKHKIKMKPTFHEDI---AAFPFGNPJHNTMT 174
 Db 115 HVNIAVLPLVYTYGKEGG---HKLNRFEGKTRVIMILNDYRAIHYITGTF----- 164
 QY 175 MRSYIDGLRYDTERMAEDYQFWYDVSFKLGRLAYPEALVKYRLHANOVSXKSHYRQ 234
 Db 165 -RRHTLFTSTVLEF-----ESIKWEEDA-----IFFNQLLKEKRYGAVAEKGYFKR 211
 QY 235 ELAIOIOKTA-----RNFLOSMGKRT-----RFDLSL-----EYR 264
 Db 212 KEQDSLVDRSFNNKRYTYLLNECYMTLMDSFNKYDVLVLOFLYHKLFLYPNKR 271
 QY 265 QTKAAAYELPKKDLPEDEFARARFL-YQCFKRTDTPPSGAMLPADGRMRRLFTLROY 323
 Db 272 DVKSYVLDQGGQGVVDFIVAKRTIDQFIKEQDMP-----MYT 311
 QY 324 FGILRLIKNRQA 337
 Db 312 KEFMFHLKENTEA 325

RESULT 9
 GGAB_BACSU STANDARD; PRT; 900 AA.
 ID GGAB_BACSU
 AC P46918;
 DT 01-NOV-1995 (Rel. 32, Created)
 DT 01-NOV-1995 (Rel. 32, last sequence update)
 DT 15-JUN-2002 (Rel. 41, last annotation update)
 DE Minor teichoic acids biosynthesis protein ggaB.
 GN GGA.
 OS Bacillus subtilis.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
 OX NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RL Submitted (Aug-1994) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,

[illegible]

Oy	270	AVEL---	-EKDPEDEPERARREFTCK	295
		:	: :	
Dd	292	IREVLYSTIDDDVITETKSVSHFYTHALK	320	
		:	: :	
RESULT	10			
ID	YG95_HAEIN	STANDARD;	PRT;	267 AA.
AC	Q48215:005081;			
DT	16-OCT-2001 (Rel. 40, Created)			
DT	16-OCT-2001 (Rel. 40, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Putative glycosyl transferase Hii1695 (EC 2.-.-.-).			
CN	Hii1695			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
OC	Haemophilus.			
OX	NCBL_TaxID=727;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=A2;			
RA	McLaughlin R., Abu Kwaik Y., Young R., Spicola S., Apicella M.,			
RT	"Characterization and sequence of the lsg locus from Haemophilus			
RL	influenzae."			
RM	Submitted (JUN-1992) to the EMBL/GenBank/DDBJ databases.			
RP	[2]			
RC	SEQUENCE FROM N.A.			
KX	STRAIN=Rd / KW20 / ATCC 51907;			
RA	MEDLINE=95350630; PubMed=7542800;			
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,			
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,			
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,			
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,			
RA	Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,			
RA	Utterback T.R., Hanna M.C., Nguyen D.T., Sauder D.M., Brandon R.C.,			
RA	Fite L.D., Fritchman J.L., Furumann J.L., Georgiagen N.S.M.,			
RA	Gnaniem C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,			
RA	Venter J.C.;			
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae			
RL	Rd.";			
RL	Science 269:496-512(1995).			
CC	-I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (see http://www.isb-sib.ch/announce/)			
CC	or send an email to license@isb-sib.ch .			
CC	-----			
DR	EMBL: M94855; AAA24983.1; .			
DR	EMBL: U32842; AAC23341.1; .			
DR	TIGR: Hii1695; .			
DR	InterPro: IPR001173; Glycos_transf_2.			
DR	Pfam: PF00535; Glycos_transf_2; 1.			
KW	Hypothetical protein; Transferase; Glycosyltransferase;			
KM	Complete proteome.			
FT	CONFLICT 26 V -> G (IN REF. 1).			
FT	CONFLICT 46 D -> E (IN REF. 1).			
FT	CONFLICT 49 F -> S (IN REF. 1).			
SO	SEQUENCE 267 AA; 30770 MW; A2FI1A0532737D8C3 CRC64;			
	Query Match	8.0%; Score 146; DB 1; Length 267;		
	Best Local Similarity	24.3%; Pred. No. 7.2e-05;		
	Matches 74; Conservative 47; Mismatches 125; Indels 54; Gaps 11;			
Oy	6	SYLVIAAYNE--KYFAQSIAAAVNQTRNLIDLIIYDDGSTDGTTLA-IADFOKRDSRIKI	62	
	:: :	: :	:	
Dd	4	SYLMALYIKENQFLRECESLVAOTRQADELVIVFDGVVPDLDEFVYTEFTK-LPLKL	62	
	:	:	:	
Oy	63	LAOANSGLIPLSLNTGLDELAKSGGGGGEYIARTDAADDIASFGWIETKIVEMEKORSIITA	122	

```

Db      63  VLLPQNRGIGKALNGLHC-----DYDWVPRMDTDICVDPREKQVAFIEQPESTII 116
Oy      123  MGAMLEVLSEEDNGRLRHHKHKIMKPTRHEDIAFFPPGNGIHNNTMTMRSDVDG 182
Db      117  FGG--QLEFEGNNVDIYAIVR-----NVPTSAOIEIKFTQRCRPFNMHTVAYQKSAVIN 168
Oy      183  GLRYDTERDMAEDYQFWYDVSKLG-RLAYYPEALVKYRLHANOVSSKSHVROH----- 234
Db      169  CGGYE---DLGEDYLYLTKIYVAGLYMANLPDILYVAVGNGVRRGVCNKAEMRLF 225
Oy      235  --ETAGIQKARNDPLQSMGFKTFFDSLEYRQTKAAAYELPEKDLPEDEFERARPLYQ 292
Db      226  KUKYRLGIQGLSLGSLFTEALRGSRLL-----LPTSLDKK-----LYQ 262
Oy      293  CFKR 296
Db      263  TFLR 266

RESULT 11
EXOO_RHIME STANDARD: PRT: 348 AA.
AC P33697;
DC 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Succinoglycan biosynthesis protein exoo (EC 2.-.-.-).
GN EXOO OR Rel084 OR SWS20959.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Bacteria; PSymb (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=94042869; PubMed=8226645;
RA Glucksmann M.A., Reuber T.L., Walker G.C.;
RT "Family of glycosyl transferases needed for the synthesis of
  RT succinoglycan by Rhizobium meliloti.";
RL J. Bacteriol. 175:7033-7044(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / SU47;
RX MEDLINE=94067019; PubMed=8246891;
RA Becker A., Kleickmann A., Keller M., Arnold W., Puehler A.;
RT "Identification and analysis of the Rhizobium meliloti exoAMONP genes
  RT involved in exopolysaccharide biosynthesis and mapping of promoters
  RT located on the exoH1AMONP fragment.";
RL Mol. Gen. Genet. 241:367-379(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=21396508; PubMed=11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
  RA Vorhoeft F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
  RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
  RT fixing endosymbiont Sinorhizobium meliloti.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -1- FUNCTION: GLYCOSYLTRANSFERASE REQUIRED FOR THE SYNTHESIS OF
  CC SUCCINOGLYCAN (EPS I). NEEDED FOR THE ADDITION OF THE FIFTH SUGAR
  CC (GLUCOSE), CATALYZES THE FORMATION OF A BETA-1,6 LINKAGE BETWEEN
  CC THE FOURTH AND FIFTH SUGAR.
CC -1- PATHWAY: Exopolysaccharide biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
  CC frameshift in position 208.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
  CC between the Swiss Institute of Bioinformatics and the EMBL outstation -

```

```

CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
  CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; L20758; AAA16044.1; ALT_FRAME.
DR EMBL; 222636; CAA80347.1; -.
DR EMBL; AL603645; CACA9484.1; -.
DR PIR; C49348; C49348.
DR PIR; S37029; S37029.
DR PIR; S3958; S3958.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycosyltransferase; Exopolysaccharide synthesis;
  KW Transferase; Glycosyltransferase;
  KW Plasmid; Complete proteome.
  FT DOMAIN
  FT SEQUENCE 348 AA; 38131 MW; 770B43782F785579 CRC64;
  SQ

Query Match 8.0%; Score 146; DB 1; Length 348;
Best Local Similarity 23.7%; Pred. No. 0.0001;
Matches 81; Conservative 51; Mismatches 144; Indels 66; Gaps 14;

Oy      3  PLVSVLICAVNYEKYFAOSLAAVNQWTRNMDILIVDDGSTDGTALAKDFQKRDSRIKI 62
Db      10  PVTEVVAAYNSADITVRAIESALAQEGVTEVVVVDOSADATFALVAIP--DPRVRL 67
Oy      63  LQAQNSGLIPSLNTGLDELAKSGGGGGEYIARTADDIASGGMTEKYEKMRKRSITA 122
Db      68  IALDRNRGGGARRNNGI-----GAARGFWIAVLSDSDPRDRRLRMTERADAGAOTA 121
Oy      123  MGAMLEVLSEEDNGRLRHHKHKIMKPTRHEDIAF-----PPGNGPIHN 170
Db      122  VDN-LDVS--LDGRSL-RMSEAEIARLP--QTLPAFIESNVLFRRSHNNGYKPIFE 175
Oy      171  NTMIRRSYIDGLRYDTERDMAEDYQFWYDVSKL--GRLAYYPEALVKYRLHANOVSS 227
Db      176  -----RRFLENGQLRFDEALRIGEDYILL--ASALACGGRCVERSA--GYIYHIREGSI 226
Oy      228  KISVQNHETAGQIQKARTARNDPLQSMGFKTFFDSLEYRQTKAAAYELPEKDLPEDEFERAR 287
Db      227  SKVLRIDHIDAMI--AADEAFLLRRYALDGLAQKMRRRRG-----FREAR 270
Oy      288  RFLYOCFKRTDPPSGAMLDFAAD-----GRMRRL 317
Db      271  SFLVVEQLKRSKSLAGALKATLADFPALRHLSMPTIARLRL 312

RESULT 12
YWDF_BACSU STANDARD: PRT: 268 AA.
AC P39614;
DC 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DE 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase ywdf (EC 2.-.-.-).
GN YWDF OR IPA-56D
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaut W., Coudart M.P., Gonzales W.,
  RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
  RA Presecan E., Santana M., Schneider E., Schweizer J., Vertes A.,
  RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
  RT kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
RN [2]
RP SEQUENCE FROM N.A.

```

RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Priescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambuit R., Wedler H., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT subdilis".
 RT Nature 390:249-256(1997).
 CC -i- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-sib.ch).
 CC
 CC EMBL: X73124; CAA51612.1;
 CC EMBL: 299123; CAB5824.1;
 CC PIR: S39711; S39711.
 CC Subtilist: Bg10602; ywdf.
 CC Interpro: IPR001173; Glycos_transf_2.
 CC Pfam: PF00535; Glycos_transf_2; 1.
 CC Hypothetical protein; Transferase; Glycosyltransferase;
 CC Complete proteome.
 CC KW
 CC SEQUENCE 268 AA; 30616 MW; D8428F7016EC9B3 CRC64;

OY 219 -- RLHANOVSSKRSV 231
 : : : : :
 Db 214 KRQYLDRLSEKHL 228
 RESULT 13
 ID SPSS_BACSU STANDARD; PRT; 256 AA.
 AC SPSS_BACSU
 DT 01-FEB-1995 (Rel. 31, Created)
 DT 01-FEB-1995 (Rel. 31, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Spore coat polysaccharide biosynthesis protein spsa.
 GN SPSS OR IPA-63D.
 OS *Bacillus subtilis*.
 OC Bacteria; Firmicutes; Bacillales; Bacillaceae; *Bacillus*.
 NCBI_TaxID=1423;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=95020537; PubMed=7934828;
 RA Glaser P., Kunst F., Arnaut M., Coudart M.P., Gonzales W.,
 RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Moszer I.,
 RA Priescan E., Santana M., Schneider E., Schweizer J., Verres A.,
 RA Rapoport G., Danchin A.;
 RT "Bacillus subtilis genome project: cloning and sequencing of the 97
 RT kb region from 325 degrees to 333 degrees".
 RL Mol. Microbiol. 10:371-384(1993).
 RP SEQUENCE FROM N.A.
 RC STRAIN=168;
 RX MEDLINE=98044033; PubMed=9384377;
 RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
 RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borchert S.,
 RA Bortiss R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
 RA Brouillet S., Brusch C.V., Caldwell B., Capuano V., Carter N.M.,
 RA Choi S.K., Codani J.J., Conneron I.F., Cummings N.J., Daniel R.A.,
 RA Dentzot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
 RA Entian K.D., Errington J., Fabret C., Ferrari E., Foulger D.,
 RA Fritz C., Fujita Y., Fuma S., Galizzi A., Galleron N.,
 RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
 RA Guisepi G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
 RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
 RA Joris B., Karamata D., Kasahara Y., Klaerr-Blanchard M., Klein C.,
 RA Kobayashi Y., Koetter P., Koningsstein G., Krogh S., Kumano M.,
 RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
 RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
 RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
 RA Noone D., O'Reilly M., Ogawa K., Ogawa A., Oudega B., Park S.H.,
 RA Parro V., Pohl T.M., Portetelle D., Porwollik S., Prescott A.M.,
 RA Priescan E., Pujic P., Purnelle B., Rapoport G., Rey M., Reynolds S.,
 RA Rieger M., Rivolta C., Rocha E., Roche B., Rose M., Sadate Y.,
 RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
 RA Sekiguchi J., Sekowska A., Seror S.J., Serror P., Shin B.S., Soldo B.,
 RA Sorokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
 RA Takeuchi M., Takakoshi A., Tanaka T., Terpstra P., Tognoni A.,
 RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassarotti A.,
 RA Viari A., Wambuit R., Wedler H., Wedler H., Weitzenecker T.,
 RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
 RA Yoshida K., Yoshikawa H.F., Zumbstein E., Yoshikawa H., Danchin A.,
 RT "The complete genome sequence of the Gram-positive bacterium *Bacillus*
 RT subdilis".
 RT Nature 390:249-256(1997).
 RN [3]
 RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).
 RX MEDLINE=99280694; PubMed=10350455;
 RA Charnock S.J., Davies G.J.;
 RT "Structure of the nucleotide-diphospho-sugar transferase, Spsa from
 RT *Bacillus subtilis*, in native and nucleotide-complexed forms.";
 CC Biochemistry 38:6380-6385(1999).
 CC -i- FUNCTION: GLYCOSYLTRANSFERASE IMPLICATED IN THE SYNTHESIS OF THE
 CC SPORE COAT.
 CC -i- PATHWAY: Spore coat polysaccharide biosynthesis.

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: X73123; CA51619.1; -
 DR EMBL: 299123; CAB15817.1; -
 DR PDB: 1Q68; 21-APR-00.
 DR PDB: 1Q68; 04-MAY-00.
 DR PDB: 1Q68; 04-MAY-00.
 DR Subtilisin: BG10609; spsa.
 DR Interpro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_1.
 DR Transferase; Glycosyltransferase; 3D-structure; Complete proteome.
 KW DISULFID 155 243
 FT SEQUENCE 256 AA; 30184 MW; C20EA9627F5D536B CRC64;
 SQ
 Query Match 7.8%; Score 143; DB 1; Length 256;
 Best Local Similarity 22.9%; Pred. No. 0.00012;
 Matches 62; Conservative 66; Mismatches 87; Indels 56; Gaps 13;
 QY 3 PLVSVLICAYNEKYPASOIAVAVNQTWRNLDILYDGSSTGTLIAADFOKRSRIKI 62
 DB 2 PAVSVMTSTYNSDYAKAISSTLSOTFSDFLPTMDNSNETLWVIRPF-LNDRKRVF 60
 QY 63 LAQONSGI-----IPSLNIGDELAKSGGGGGEIARTDADDIASPGWIEKIVE 113
 DB 61 Y-QSDISGVKERTKRYVALINQAI-EMAE-----GEITYATDNTNIMPDLLKVMRE 113
 QY 114 ME-KRSTIAGCAMLEVISEEKDGNRLARHHKHKIKKPRRHEDIAAFPPGNG--IH 169
 DB 114 LTHPKAVIYASKTYHLENND-----IYKETYAP--AAQVTWNAACAD 158
 QY 170 NNTMIRRSVIDGLRYPDERDPAE-----DYQFWYDVSGLRGLAYPEALVYRHL 221
 DB 159 HCSVMHRYVLE-KVKEKGSYWDSPARYIGDARFFRVNH-----FYPRPLDEILD 212
 QY 222 ANOVSSKHSVRQHEIAQGIQKTARNDELQSM 252
 DB 213 LNVITDQ-----SIHQFLFELEKNEFVRL 237
 RESULT 14
 YCDQ_ECOLI
 ID YCDQ_ECOLI STANDARD; PRT; 441 AA.
 AC P75905;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Hypothetical protein ycdQ.
 GN YCDQ OR B1022.
 OS Escherichia coli.
 OC Escherichia coli.
 OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
 OC Escherichia.
 OC NCBI_TaxID=562;
 RX NCBI_TaxID=562;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12 / MG1655;
 RX MEDLINE=97426617; PubMed=9278503;
 RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
 RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
 RA Grou B., Shao Y.,
 RA Mau G., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
 RT "The complete genome sequence of Escherichia coli K-12."
 RL Science 277:1453-1474(1997).
 RN [2]
 RP SEQUENCE FROM N.A.
 RC STRAIN-K12;
 RX STRAIN-K12;

RX MEDLINE=97061202; PubMed=8905232;
 RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
 RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
 RA Kimura S., Kitagawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
 RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
 RA Samped G., Seki Y., Tagami H., Takemoto H., Wada C., Yamamoto Y.,
 RA Yano M., Horiuchi T.;
 RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
 RT corresponding to the 12.7-28.0 min region on the linkage map."
 RL DNA Res. 3:137-155(1996).
 CC -1- SUBCELLULAR LOCATION: Integral membrane protein (Potential).
 CC -1- SIMILARITY: TO Y.PESTIS HMS LOCUS PROTEIN HMSR AND TO
 CC S.EPIDERMIDIS ICA.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 DR EMBL: AE000204; AAC74107.1; -
 DR EMBL: D90739; BAA35803.1; -
 DR EcoGene; EGI3863; ycdQ.
 DR Interpro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_1.
 DR Hypothetical protein; Transmembrane; Complete proteome.
 KW TRANSMEM 5 25
 FT TRANSMEM 33 53
 FT TRANSMEM 331 351
 FT TRANSMEM 363 383
 FT TRANSMEM 395 415
 FT SEQUENCE 441 AA; 50765 MW; 04F5A53D72FEBABB CRC64;
 SQ
 Query Match 7.4%; Score 135; DB 1; Length 441;
 Best Local Similarity 23.9%; Pred. No. 0.0011;
 Matches 57; Conservative 35; Mismatches 86; Indels 60; Gaps 9;
 QY 3 PLVSVLICAYNEKYPASOIAVAVNQTWRNLDILYDGSSTGTLIAADFOKRSRIKI 62
 DB 75 PSTSIITPCFNEKNEKVEITHALAQRYENIEVIAVNDSTDKTRAILDRMAQIPHLRV 134
 QY 63 LAQONSGIIPSLNIGDELAKSGGGGGEIARTDAD-----IASPGWIEKIVE 113
 DB 135 IHLAQNGKAIKTKGA-AAAKS-----EYLCIDGDALLDRAAYIYEPMLYNRVGA 188
 QY 114 MEKRSI-----IANGAMLEVISEEKDGNRLARHHKHKIKKPRRHEDIAAFPPF 164
 DB 189 VTGNPRIRTRSTLVGRIGYGEYSIIIGLIKRTQRI-----YGNVF--TVSGVIAAF-- 237
 QY 165 GNPINHTMIMRSYV-----DGLRYPDERDPAEDYQFWYDVSGLRGLAYPEAL 215
 DB 238 -----RRSALAEVYSDMDITDIDISWKLQINQW-----TIFYEPRL 277
 RESULT 15
 Y060_MYCPN
 ID Y060_MYCPN STANDARD; PRT; 299 AA.
 AC P75042;
 DT 01-NOV-1997 (Rel. 35, Created)
 DT 01-NOV-1997 (Rel. 35, Last sequence update)
 DT 16-OCT-2001 (Rel. 40, Last annotation update)
 DE Putative glycosyl transferase MG060 homolog (EC 2.-.-.-)
 GN (D09_ort299).
 GN MPN075 OR MP080.
 OS Mycoplasma pneumoniae.
 OC Bacteria; Firmicutes; Mollicutes; Mycoplasmataceae; Mycoplasma.
 OC NCBI_TaxID=2104;
 RX NCBI_TaxID=2104;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=ATCC 29342 / M129;
 RX MEDLINE=97105885; PubMed=8948633;

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 : Search time 29.8926 Seconds
(without alignments)
2398.732 Million cell updates/sec

Title: US-10-007-267-3

Perfect score: 1823

Sequence: 1 MQLPLSVLICAVNVEKYEFAQ.....RLIKNRQARSDSAGKEGEI 348

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

- 1: SPREMBL_21.*
- 2: sp_archaea.*
- 3: sp_bacteria.*
- 4: sp_fungi.*
- 5: sp_human.*
- 6: sp_invertebrate.*
- 7: sp_mammal.*
- 8: sp_minc.*
- 9: sp_organelle.*
- 10: sp_phage.*
- 11: sp_plant.*
- 12: sp_protent.*
- 13: sp_virus.*
- 14: sp_vertebrate.*
- 15: sp_unclassified.*
- 16: sp_rvirus.*
- 17: sp_bacteriap.*
- 17: sp_archaeap.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1823	100.0	348	2	Q50946
2	1788	98.1	346	2	Q9RGN2
3	1785	97.9	362	2	Q50951
4	1773	97.3	346	2	Q93PS0
5	1712	93.9	346	16	Q9JW66
6	1657.5	90.9	322	2	Q9L8F0
7	1651	90.6	321	2	Q9L8E9
8	1621.5	88.9	349	16	Q9JX06
9	1617.5	88.7	333	2	Q51115
10	1609	88.3	348	2	Q9EVD5
11	1118	61.3	221	2	Q93EK6
12	1049	57.5	337	2	Q50949
13	666.5	36.6	337	16	Q9CIE9
14	412.5	22.6	337	2	P74947
15	410	22.5	337	2	Q34234
16	373	20.5	172	2	Q9RGW9

17	346	19.0	149	2	P66946	P66946 neisseria m
18	327	17.9	367	16	Q92V61	Q92V61 rhizobium m
19	318	17.4	80	2	Q9K2R3	Q9K2R3 neisseria g
20	312	17.1	368	16	Q98JH2	Q98JH2 rhizobium 1
21	303	16.6	117	2	P66944	P66944 neisseria m
22	300	16.5	133	2	Q93EK8	Q93EK8 neisseria m
23	293.5	16.1	333	16	Q97H38	Q97H38 clostridium
24	279.5	15.3	340	2	Q87159	Q87159 vibrio chol
25	276	15.1	321	16	Q8YSW2	Q8YSW2 anabena sp
26	274	15.0	388	2	Q56869	Q56869 yerstinla en
27	272	14.9	333	16	Q97H40	Q97H40 clostridium
28	271	14.9	336	16	Q97H39	Q97H39 clostridium
29	267.5	14.7	314	2	Q9LA88	Q9LA88 aeromonas h
30	258	14.2	271	16	Q9K615	Q9K615 bacillus ha
31	252.5	13.9	343	16	Q8YUP7	Q8YUP7 anabena sp
32	247	13.5	732	16	Q9K650	Q9K650 bacillus ha
33	245.5	13.5	318	16	Q8YSL3	Q8YSL3 anabena sp
34	236.5	13.0	318	16	Q8YSL4	Q8YSL4 anabena sp
35	235	12.9	333	16	Q8YWS0	Q8YWS0 anabena sp
36	234.5	12.9	354	2	Q9AH91	Q9AH91 streptococc
37	230.5	12.6	1013	16	Q8YDZ1	Q8YDZ1 anabena sp
38	229.5	12.6	337	16	Q8YD6	Q8YD6 anabena sp
39	229.5	12.6	696	16	Q97P72	Q97P72 streptococc
40	228.5	12.5	306	2	Q9RP62	Q9RP62 escherichia
41	228	12.5	257	2	Q93CQ9	Q93CQ9 shigella bo
42	228	12.5	278	16	P71054	P71054 bacillus su
43	227	12.5	279	2	Q9RG44	Q9RG44 streptococc
44	225	12.3	298	17	Q9UZ16	Q9UZ16 pyrococcus
45	224	12.3	301	2	Q9S410	Q9S410 leptospira

ALIGNMENTS

RESULT 1	
Q50946	PRELIMINARY; PRT; 348 AA.
AC	Q50946; 01-NOV-1996 (TREMBLrel. 01, Created)
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT	01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE	Glycosyl transferase.
GN	LGTA.
OS	Neisseria gonorrhoeae.
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX	NCBI_TaxID=485;
RN	[1]
RP	SEQUENCE FROM N.A.
RC	STRAIN=F62;
RX	MEDLINE=95053752; PubMed=7964493;
RA	Gotschlich E.C.;
RT	"Genetic locus for the biosynthesis of the variable portion of
RT	Neisseria gonorrhoeae lipooligosaccharide.";
RL	J. Exp. Med. 180:2181-2190(1994).
DR	EMBL; U14554; AAA68009.1;
DR	InterPro: IPR001173; Glycos_transf_2.
DR	Pfam: PF00535; Glycos_transf_2; 1.
KW	Transferase.
SO	SEQUENCE 348 AA; 39975 MW; FCBPF2FACD84C72A CRC64;
Query Match	100.0%; Score 1823; DB 2; Length 348;
Best Local Similarity	100.0%; Pred. No. 1.5e-148;
Matches 348;	Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY	1 MQLPLSVLICAVNVEKYEFAQSLAAVYVQTRNLDTIIVDGSSTDTGLAIKQFQRDSRI 60
DB	1 MQLPLSVLICAVNVEKYEFAQSLAAVYVQTRNLDTIIVDGSSTDTGLAIKQFQRDSRI 60
QY	61 KTLAAQNSGLIPSLNIGIDELAKSGGGGEYIARRDADDIASPGYIEYIVGEMEDRSI 120
DB	61 KTLAAQNSGLIPSLNIGIDELAKSGGGGEYIARRDADDIASPGYIEYIVGEMEDRSI 120
QY	121 IAMGAWLEVLSEKDGKGNRLARRHKKGKIKKPTRHEDIAAPFGNPIHNNMTIMRSYI 180

```
|||||
DB 121 IAMGAWLEVLSEKQGNRLARHHKHKIWKKPTRHEDIAAFPEFGNPINNNTMIMRSYI 180
QY 181 DQGLRYDTERDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHVROHEIAGOI 240
DB 181 DQGLRYDTERDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHVROHEIAGOI 240
QY 241 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
DB 241 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
QY 301 PSGAWLDFPAADGRMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
DB 301 PSGAWLDFPAADGRMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
```

```
RESULT 2
Q9RGN2 PRELIMINARY; PRT; 346 AA.
ID Q9RGN2
AC Q9RGN2;
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
DE Glycosyltransferase.
GN LGTA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1291;
RX MEDLINE=20305049; PubMed=10844691;
RA Harvey H.A., Porat N., Campbell C.A., Jennings M., Gibson B.W.,
  Phillips N.J., Apicella M.A., Balke M.S.;
  "Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein
  receptor on human sperm.";
  Mol. Microbiol. 36:1059-1070(2000).
RL EMBL, AF121135; AAF14359.1; -.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Transferase.
SQ SEQUENCE 346 AA; 39730 MW; A7E9434C4635815D CRC64;
```

```
Query Match 98.1%; Score 1788; DB 2; Length 346;
Best Local Similarity 98.6%; Pred. No. 1.5e-145;
Matches 343; Conservative 2; Mismatches 1; Indels 2; Gaps 1;
```

```
QY 1 MOPVSVLICAVNVEKYFAQSLAAVYNQWNRMLDILIVDGSIDGTLATAKDFOKRDSRI 60
DB 1 MOPVSVLICAVNVEKYFAQSLAAVYNQWNRMLDILIVDGSIDGTLATAKDFOKRDSRI 60
QY 61 KIQAQONSGILPSLNINGDELAKSGGGEYIARTDADDIASPGVIEKIVGEMEKDRSI 120
DB 61 KIQAQONSGILPSLNINGDELAKSGGGEYIARTDADDIASPGVIEKIVGEMEKDRSI 118
QY 121 IAMGAWLEVLSEKQGNRLARHHKHKIWKKPTRHEDIAAFPEFGNPINNNTMIMRSYI 180
DB 121 IAMGAWLEVLSEKQGNRLARHHKHKIWKKPTRHEDIAAFPEFGNPINNNTMIMRSYI 178
QY 181 DQGLRYDTERDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHVROHEIAGOI 240
DB 181 DQGLRYDTERDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHVROHEIAGOI 238
QY 241 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
DB 241 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 298
QY 301 PSGAWLDFPAADGRMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
DB 299 PSGAWLDFPAADGRMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 346
```

RESULT 3

```
Q50951
ID Q50951 PRELIMINARY; PRT; 362 AA.
AC Q50951;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, last annotation update)
DE Glycosyltransferase.
GN LS12.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1291;
RX MEDLINE=96099316; PubMed=8522539;
RA Danaher R.J., Levin J.C., Arking D., Burch C.L., Sandlin R.,
  Stein D.C.;
  "Genetic basis of Neisseria gonorrhoeae lipooligosaccharide antigenic
  variation.";
  J. Bacteriol. 177:7275-7279(1995).
RL EMBL, U15992; AAA92074.1; -.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Transferase.
SQ SEQUENCE 362 AA; 41623 MW; ED67B468C9E3C3A CRC64;
```

```
Query Match 97.9%; Score 1785; DB 2; Length 362;
Best Local Similarity 98.3%; Pred. No. 3e-145;
Matches 342; Conservative 3; Mismatches 1; Indels 2; Gaps 1;
```

```
QY 1 MOPVSVLICAVNVEKYFAQSLAAVYNQWNRMLDILIVDGSIDGTLATAKDFOKRDSRI 60
DB 17 MOPVSVLICAVNVEKYFAQSLAAVYNQWNRMLDILIVDGSIDGTLATAKDFOKRDSRI 76
QY 61 KIQAQONSGILPSLNINGDELAKSGGGEYIARTDADDIASPGVIEKIVGEMEKDRSI 120
DB 77 KIQAQONSGILPSLNINGDELAKSGGGEYIARTDADDIASPGVIEKIVGEMEKDRSI 134
QY 121 IAMGAWLEVLSEKQGNRLARHHKHKIWKKPTRHEDIAAFPEFGNPINNNTMIMRSYI 180
DB 135 IAMGAWLEVLSEKQGNRLARHHKHKIWKKPTRHEDIAAFPEFGNPINNNTMIMRSYI 194
QY 181 DQGLRYDTERDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHVROHEIAGOI 240
DB 195 DQGLRYDTERDMAEDYQFYVDVSKLGRLAYYPEALVKYRLHANOVSSKSHVROHEIAGOI 254
QY 241 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 300
DB 255 OKTARNDFLOSMGFRTFRPSLEYRQTKAAAYELPEKDLPEEDFERARRFLYOCFKRTDTP 314
QY 301 PSGAWLDFPAADGRMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 348
DB 315 PSGAWLDFPAADGRMRRLFTLRQYFGILYRLIKNRQARSDSAGKEOEI 362
```

```
RESULT 4
Q93PS0 PRELIMINARY; PRT; 346 AA.
ID Q93PS0
AC Q93PS0;
DT 01-DEC-2001 (TREMBLrel. 19, Created)
DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, last annotation update)
DE Glycosyl transferase Lgta.
GN LGTA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PID2;
RA Tong Y., Arking D., Reinhold V., Stein D.C.;
  "Characterization of lipooligosaccharide structures found in Neisseria
  gonorrhoeae PID2.";
```

RL Submitted (OCT-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF313394; AAK70338.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Transferrase.
 SO SEQUENCE 346 AA; 39727 MW; 37C680CD3B14984 CRC64;

Query Match 97.3%; Score 1773; DB 2; Length 346;
 Best Local Similarity 97.4%; Pred. No. 3e-144;
 Matches 339; Conservative 5; Mismatches 2; Indels 2; Gaps 1;

QY 1 MNPVSVLICAVNVEKYPQSLAAVYVQWRLDILYVDDGSTDGLAIKPFQKRDSTI 60
 DB 1 MNPVSVLICAVNVEKYPQSLAAVYVQWRLDILYVDDGSTDGLAIKPFQKRDSTI 60
 QY 61 KILAAQNSGLIPSLINIGDELAKSGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
 DB 61 KILAAQNSGLIPSLINIGDELAKSGGGG-YIARTDADDIAPGMIEKIVGEMEKDRSI 118
 QY 121 IAMGAMLEVLSEKDGKGNRLARRHHKGIWKKPTRHEDIAAEPFGNPINHTMIMRSYI 180
 DB 121 IAMGAMLEVLSEKDGKGNRLARRHHKGIWKKPTRHEDIAAEPFGNPINHTMIMRSYI 178
 QY 119 IAMGAMLEVLSEKDGKGNRLARRHHKGIWKKPTRHEDIAAEPFGNPINHTMIMRSYI 178
 DB 119 IAMGAMLEVLSEKDGKGNRLARRHHKGIWKKPTRHEDIAAEPFGNPINHTMIMRSYI 178
 QY 181 DGLRYDTERDMAEDYQFVYDVSKLGRLAYYPEALVYKRLHANOVSSKHSVQHETAGCI 240
 DB 179 DGLRYDTERDMAEDYQFVYDVSKLGRLAYYPEALVYKRLHANOVSSKHSVQHETAGCI 238
 QY 241 OKTARNDFLOSGKFTRPDSLEYROTAKAAYELPEKDLPEEFERARRLYOCFKRTDP 300
 DB 239 OKTARNDFLOSGKFTRPDSLEYROTAKAAYELPEKDLPEEFERARRLYOCFKRTDP 298
 QY 301 PSGAMLDFAADGRMRRLFTLRQYFGILYRLIKNRQARSDSAGKDEI 348
 DB 299 PSGAMLDFAADGRMRRLFTLRQYFGILYRLIKNRQARSDSAGKDEI 346

RESULT 5

QYJW66 PRELIMINARY; PRT; 346 AA.
 AC QJW66;
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-DEC-2001 (TREMblrel. 19, last annotation update)
 DE Lacto-N-neotetraose blosynthesis glycosyl transferase.
 GN LGTA OR NMA0524.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxId=65699;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / SEROGROUP A / SEROTYPE 4A;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parthall J., Achman M., James K.D., Bentley S.D., Churcher C.,
 Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 Davies R.M., Davis P., Devlin K., Feltwell T., Hamlin N., Holtroyd S.,
 Jajels K., Leather S., Moule S., Mungall K., Quail M.A.,
 Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 Whitehead S., Spratt B.G., Barrrell B.G.,
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 meningitidis 22491".
 RL Nature 404:502-506(2000).
 DR EMBL: AL162753; CAB83816.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 KW Complete proteome.
 SO SEQUENCE 346 AA; 40158 MW; 95C89CBA01EFA9 CRC64;

Query Match 93.9%; Score 1712; DB 16; Length 346;
 Best Local Similarity 94.3%; Pred. No. 5.3e-139;
 Matches 328; Conservative 10; Mismatches 8; Indels 2; Gaps 1;

QY 1 MNPVSVLICAVNVEKYPQSLAAVYVQWRLDILYVDDGSTDGLAIKPFQKRDSTI 60
 DB 1 MNPVSVLICAVNVEKYPQSLAAVYVQWRLDILYVDDGSTDGLAIKPFQKRDSTI 60

DB 1 MNPVSVLICAVNVEKYPQSLATVYVQWRLDILYVDDGSTDGLAIKPFQKRDSTI 60
 QY 61 KILAAQNSGLIPSLINIGDELAKSGGGGEYIARTDADDIASPGMIKIVGEMEKDRSI 120
 DB 61 KILAAQNSGLIPSLINIGDELAKS--GMGEYIARTDADDIAPGMIEKIVGEMEKDRSI 118
 QY 121 IAMGAMLEVLSEKDGKGNRLARRHHKGIWKKPTRHEDIAAEPFGNPINHTMIMRSYI 180
 DB 119 IAMGAMLEVLSEKDGKGNRLARRHHKGIWKKPTRHEDIAAEPFGNPINHTMIMRSYI 178
 QY 181 DGLRYDTERDMAEDYQFVYDVSKLGRLAYYPEALVYKRLHANOVSSKHSVQHETAGCI 240
 DB 179 DGLRYDTERDMAEDYQFVYDVSKLGRLAYYPEALVYKRLHANOVSSKHSVQHETAGCI 238
 QY 241 OKTARNDFLOSGKFTRPDSLEYROTAKAAYELPEKDLPEEFERARRLYOCFKRTDP 300
 DB 239 OKTARNDFLOSGKFTRPDSLEYROTAKAAYELPEKDLPEEFERARRLYOCFKRTDP 298
 QY 301 PSGAMLDFAADGRMRRLFTLRQYFGILYRLIKNRQARSDSAGKDEI 348
 DB 299 PAGAMLDFAADGRMRRLFTLRQYFGILYRLIKNRQARSDSAGKDEI 346

RESULT 6

QYJ8F0 PRELIMINARY; PRT; 322 AA.
 AC QJL8F0;
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-OCT-2000 (TREMblrel. 15, last sequence update)
 DT 01-JUN-2001 (TREMblrel. 17, last annotation update)
 DE Lgta.
 GN LGTA.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxId=485;
 RN 111
 RP SEQUENCE FROM N.A.
 RC STRAIN=F4899;
 RA Balhazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
 RT "Mutations in the lgt operon influence serum-resistance in
 gonococci".
 RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AF208056; AAF25876.1; -
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2; 1.
 SO SEQUENCE 322 AA; 37109 MW; 88FECASAD41DC2B7 CRC64;

Query Match 90.9%; Score 1657.5; DB 2; Length 322;
 Best Local Similarity 97.8%; Pred. No. 2.3e-134;
 Matches 316; Conservative 3; Mismatches 3; Indels 1; Gaps 1;

QY 26 VNOTWRNLDILYVDDGSTDGLAIKPFQKRDSTI KILAAQNSGLIPSLINIGDELAKS 85
 DB 1 MNPVSVLICAVNVEKYPQSLATVYVQWRLDILYVDDGSTDGLAIKPFQKRDSTI KILAAQNSGLIPSLINIGDELAKS 60
 QY 86 GGGGEYIARTDADDIASPGMIKIVGEMEKDRSI IAMGAMLEVLSEKDGKGNRLARRHHK 145
 DB 61 -GGGGEYIARTDADDIASPGMIKIVGEMEKDRSI IAMGAMLEVLSEKDGKGNRLARRHHK 119
 QY 146 GKIWKPTRHEDIAAEPFGNPINHTMIMRSYIDGRLYDTERDMAEDYQFVYDVSKL 205
 DB 120 GAIWPKPTRHEDIAAEPFGNPINHTMIMRSYIDGRLYDTERDMAEDYQFVYDVSKL 179
 QY 206 GLRAYYPEALVYKRLHANOVSSKHSVQHETAGCI OKTARNDFLOSGKFTRPDSLEYRQ 265
 DB 180 GLRAYYPEALVYKRLHANOVSSKHSVQHETAGCI OKTARNDFLOSGKFTRPDSLEYRQ 239
 QY 266 TKAAYELPEKDLPEEFERARRLYOCFKRTDP PSGAMLDFAADGRMRRLFTLRQYFG 325
 DB 240 TKAAYELPEKDLPEEFERARRLYOCFKRTDP PSGAMLDFAADGRMRRLFTLRQYFG 299
 QY 326 ILYRLIKNRQARSDSAGKDEI 348
 DB 326 ILYRLIKNRQARSDSAGKDEI 348

DB 300 ILRLIKNRQARSAGKEOEI 322

RESULT 7

ID 0918E9 PRELIMINARY; PRT; 321 AA.

AC 0918E9;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)

DE 01-JUN-2001 (TREMBLrel. 17, last annotation update)

GN LGTA.

OS Neisseria gonorrhoeae.

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=485;

RP [1]

RC SEQUENCE FROM N.A.

RC STRAIN=SS2;

RA Balhazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;

RT "Mutations in the 1st operon influence serum-resistance in gonococci";

RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.

DR EMBL: AF208059; AAF2877.1; -

DR InterPro: IPR001173; Glycos_transf.2.

DR Pfam: PF00535; Glycos_transf.2.1.

SQ SEQUENCE 321 AA; 37052 MW; 6ACADA9A3CB738FA CRC64;

Query Match 90.6%; Score 1651; DB 2; Length 321;
Best Local Similarity 97.5%; Pred. No. 8.4e-134;
Matches 315; Conservative 3; Mismatches 3; Indels 2; Gaps 1;

QY 26 VNQTNWNLIDLIYDDGSTDTLAIANDFOKRDRIKILQAQNSGLIPSLINIGDELAKS 85
DB 1 NMQTNWNLIDLIYDDGSTDTLAIANDFOKRDRIKILQAQNSGLIPSLINIGDELAKS 60
QY 86 GGGGGYIARTDADDDIASCWIEKTYGEMEKDRSTIANGAMLEVESEKDNRLAHNRH 145
DB 61 --GGGGYIARTDADDDIASCWIEKTYGEMEKDRSTIANGAMLEVESEKDNRLAHNRH 118
QY 146 GKIKKPTHEHEDIAAFPPGPNPHNNTMTIMRSVIDGRLRYDTERWADYQFWYVSKL 205
DB 119 GAIWKPFTHEHEDIAAFPPGPNPHNNTMTIMRSVIDGRLRYDTERWADYQFWYVSKL 178
QY 206 GRLAYPEALVYKRLHANOVSSKSHVROHEIAQIQKTARNDELQSGMGEFTRPDSLEYRQ 265
DB 179 GRLAYPEALVYKRLHANOVSSKSHVROHEIAQIQKTARNDELQSGMGEFTRPDSLEYRQ 238
QY 266 TFAAAEELPEKDLPEDEFRARRELYQCFKRTDTPPSGAMLDFAADGRMRRLFTLQYRFG 325
DB 239 TFAAAEELPEKDLPEDEFRARRELYQCFKRTDTPPSGAMLDFAADGRMRRLFTLQYRFG 298
QY 326 ILRLIKNRQARSAGKEOEI 348
DB 299 ILRLIKNRQARSAGKEOEI 321

RESULT 8

Q9JX06 PRELIMINARY; PRT; 349 AA.

AC 09JX06;

DT 01-OCT-2000 (TREMBLrel. 15, Created)

DT 01-OCT-2000 (TREMBLrel. 15, last sequence update)

DE 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE Lacto-N-neotetraose biosynthesis glycosyl transferase Lgta.

GN NMB1929.

OS Neisseria meningitidis (serogroup B).

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=491;

RP [1]

RC SEQUENCE FROM N.A.

RC STRAIN=MC58 / SEROGROUP B;

RX MEDLINE=20175755; PubMed=10710307;

RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,

RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Utterback T.R., Khouri H., Qin H., Vanachavan J.,
RA Gill J., Scariato V., Maignani V., Pizzo M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappuoli R., Venter J.C.;

RT "Complete genome sequence of Neisseria meningitidis serogroup B strain MC58";

RL Science 287:1809-1815(2000).

DR EMBL: AE002541; AAF42358.1; -

DR TIGR: NMB1929; -

DR InterPro: IPR001173; Glycos_transf.2.

DR Pfam: PF00535; Glycos_transf.2.1.

KW Transferase; Complete proteome.

SQ SEQUENCE 349 AA; 40618 MW; 4B9E245C34847DC3 CRC64;

Query Match 88.9%; Score 1621.5; DB 16; Length 349;
Best Local Similarity 92.5%; Pred. No. 3.3e-131;
Matches 309; Conservative 14; Mismatches 10; Indels 1; Gaps 1;

QY 1 MQLVSVLICATNVEKFFQSLAAVYNQTRNDLIYDDGSTDTLAIANDFOKRDRI 60
DB 17 LQPLVSVLICATNVEKFFQSLAAVYNQTRNDLIYDDGSTDTLAIANDFOKRDRI 76
QY 61 KILQAQNSGLIPSLINIGDELAKSGGGGEYIARTDADDDIASCWIEKTYGEMEKDRSTI 120
DB 77 RIIAQGRNSGLIPSLINIGDELAKS-GGGGEYIARTDADDDIASCWIEKTYGEMEKDRSTI 135
QY 121 IANGAMLEVESEKDNRLARHHKGIKWKPTREHEDIAAFPPGPNPHNNTMTIMRSYI 180
DB 136 IANGAMLEVESEKDNRLARHHKGIKWKPTREHEDIAAFPPGPNPHNNTMTIMRSYI 195
QY 181 DGLRYDTERWADYQFWYVSKLGRLAYPEALVYKRLHANOVSSKSHVROHEIAQIQ 240
DB 196 DGLRYDTERWADYQFWYVSKLGRLAYPEALVYKRLHANOVSSKSHVROHEIAQIQ 255
QY 241 QKTARNDELQSGMGEFTRPDSLEYRQTKAAAYELPEKDLPEDEFRARRELYQCFKRTDTP 300
DB 256 QKTARNDELQSGMGEFTRPDSLEYRQTKAAAYELPEKDLPEDEFRARRELYQCFKRTDTP 315
QY 301 PSGAMLDFAADGRMRRLFTLQYRFGILRLIKNR 334
DB 316 PSGAMLDFAADGRMRRLFTLQYRFGILRLIKNR 349

RESULT 9

ID 051115 PRELIMINARY; PRT; 333 AA.

AC 051115;

DT 01-NOV-1996 (TREMBLrel. 01, Created)

DT 01-NOV-1996 (TREMBLrel. 01, last sequence update)

DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)

DE Glycosyl transferase.

GN LGTA.

OS Neisseria meningitidis.

OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.

OX NCBI_TaxID=487;

RP [1]

RC SEQUENCE FROM N.A.

RC STRAIN=MC58;

RX MEDLINE=96414473; PubMed=8817494;

RA Jennings M.P., Hood D., Peak I.R.A., Virji M., Moxon E.R.;

RT "Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in Neisseria meningitidis";

RT Mol. Microbiol. 18:729-740(1995).

RL EMBL: U25839; AAC44084.1; -

DR InterPro: IPR001173; Glycos_transf.2.

DR Pfam: PF00535; Glycos_transf.2.1.

SQ SEQUENCE 333 AA; 38563 MW; 49D8F6CE375387BF CRC64;

Query Match 88.7%; Score 1617.5; DB 2; Length 333;

Best Local Similarity 92.5%; Pred. No. 6.8e-131;
Matches 309; Conservative 13; Mismatches 11; Indels 1; Gaps 1;

```

QY 1 MOPVSVLICAYNEKFAOSLAIVNQTNRNLDILYVDSSTGTLIAAKPQKRSRI 60
   |||||||
Db 1 MOPVSVLICAYNEKFAOSLAIVNQTNRNLDILYVDSSTGTLIAARFEOBDR 60
   |||||||
QY 61 KIILAOQNSGLIPSLNIGLDELAKSGGGGGGYIARTDADIASPGWIEKIVGEMEKDRSI 120
   |||||
Db 61 RIILQPNRNSGLIPSLNIGLDELAKS--GGGGGYIARTDADIIAADWIEKIVGEMEKDRSI 119
   |||||
QY 121 IAMGAMLEVLSEEEKDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMTIMRSVI 180
   |||||||
Db 120 IAMGAMLEVLSEEEKDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMTIMRSVI 179
   |||||||
QY 181 DGGRLYPTERMADYOFWYDVSKLGRLAYYPEALVYKRLANOVSSKHSYRHOEIIAQGI 240
   |||||||
Db 180 DGGRLYPTERMADYOFWYDVSKLGRLAYYPEALVYKRLANOVSSKHSYRHOEIIAQGI 239
   |||||||
QY 241 OKTARNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQCFKRTDTP 300
   |||||||
Db 240 OKTARNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQCFKRTDTP 299
   |||||||
QY 301 PSGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 334
   |||||||
Db 300 PAGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 333
   |||||||

```

RESULT 10

```

Q9EVD5 ID Q9EVD5 PRELIMINARY; PRT; 348 AA.
AC Q9EVD5;
DT 01-MAR-2001 (TReMBLrel, 16, Created)
DT 01-MAR-2001 (TReMBLrel, 16, Last sequence update)
DE 01-JUN-2002 (TReMBLrel, 21, Last annotation update)
DE N-acetylglycosamine transferase.
GN LGTA.
OS Neisseria subflava.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_Taxid=28449;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21142520; Pubmed=11208792;
RA Aking D., Tong Y., Stein D.C.;
RT "Analysis of lipooligosaccharide biosynthesis in the Neisseriaceae.";
RL J. Bacteriol. 183:934-941(2001).
DR EMBL: AF240672; AAC09764.1; -.
DR HSSP: P39621; 10GQ.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KM Transferase.
SQ SEQUENCE 348 AA; 40676 MW; BC9E313E9BC0BF41 CRC64;

```

Query Match 88.3%; Score 1609; DB 2; Length 348;
Best Local Similarity 92.2%; Pred. No. 3.9e-130;
Matches 308; Conservative 13; Mismatches 11; Indels 2; Gaps 1;

```

QY 1 MOPVSVLICAYNEKFAOSLAIVNQTNRNLDILYVDSSTGTLIAAKPQKRSRI 60
   |||||||
Db 17 LOPVSVLICAYNEKFAOSLAIVNQTNRNLDILYVDSSTGTLIAARFEOBDR 76
   |||||||
QY 61 KIILAOQNSGLIPSLNIGLDELAKSGGGGGGYIARTDADIASPGWIEKIVGEMEKDRSI 120
   |||||
Db 77 RIILQPNRNSGLIPSLNIGLDELAKS--GGGGGYIARTDADIIAADWIEKIVGEMEKDRSI 134
   |||||
QY 121 IAMGAMLEVLSEEEKDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMTIMRSVI 180
   |||||||
Db 135 IAMGAMLEVLSEEEKDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMTIMRSVI 194
   |||||||
QY 181 DGGRLYPTERMADYOFWYDVSKLGRLAYYPEALVYKRLANOVSSKHSYRHOEIIAQGI 240
   |||||||
Db 195 DGGRLYPTERMADYOFWYDVSKLGRLAYYPEALVYKRLANOVSSKHSYRHOEIIAQGI 254
   |||||||

```

QY 241 OKTARNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQCFKRTDTP 300
|||||||
Db 255 OKTARNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQCFKRTDTP 314
|||||||

QY 301 PSGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 334
|||||||
Db 315 PAGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 348
|||||||

RESULT 11

```

Q93EK6 ID Q93EK6 PRELIMINARY; PRT; 221 AA.
AC Q93EK6;
DT 01-DEC-2001 (TReMBLrel, 19, Created)
DT 01-DEC-2001 (TReMBLrel, 19, Last sequence update)
DE 01-DEC-2001 (TReMBLrel, 19, Last annotation update)
DE LGTA.
GN LGTA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_Taxid=487;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=M978;
RX MEDLINE=21467954; Pubmed=11563844;
RA Zhu P., Klutch M.J., Tsai C.-M.;
RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide Expression in Two 18-Immunotype Strains of Neisseria meningitidis.";
RL FEMS Microbiol. Lett. 203:173-177(2001).
DR EMBL: AF355193; AAL12840.1; -.
SQ SEQUENCE 221 AA; 26478 MW; D1F97099B1F6D55F CRC64;

```

Query Match 91.3%; Score 1118; DB 2; Length 221;
Best Local Similarity 94.6%; Pred. No. 3.5e-88;
Matches 209; Conservative 6; Mismatches 6; Indels 0; Gaps 0;

```

QY 114 MEKDSIIAMGAMLEVLSEEEKDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMT 173
   |||||||
Db 1 MEKDSIIAMGAMLEVLSEEEKDGNRLARHHRGKIWKPTRHEDIAAFPPGPNPIHNTMT 60
   |||||||
QY 174 IMRSYVDGRLYPTERMADYOFWYDVSKLGRLAYYPEALVYKRLANOVSSKHSYRQ 233
   |||||||
Db 61 IMRSYVDGRLYPTERMADYOFWYDVSKLGRLAYYPEALVYKRLANOVSSKHSYRQ 120
   |||||||
QY 234 HEIIQGIQKTRNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQC 293
   |||||||
Db 121 HEIIQGIQKTRNDFLOSMGFRTFSDSLEYRQTKAAAYELPEKDLPEEDFERARRFLYQC 180
   |||||||
QY 294 FKRTDTPPSGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 334
   |||||||
Db 181 FKRTDTPPSGAMLDPAADGRMRRLFTLRQYFGILYRLIKNR 221
   |||||||

```

RESULT 12

```

Q50949 ID Q50949 PRELIMINARY; PRT; 337 AA.
AC Q50949;
DT 01-NOV-1996 (TReMBLrel, 01, Created)
DT 01-NOV-1996 (TReMBLrel, 01, Last sequence update)
DE 01-JUN-2001 (TReMBLrel, 17, Last annotation update)
DE Glycosyl transferase.
GN LGTD.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_Taxid=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F62;
RX MEDLINE=95053752; Pubmed=7964493;
RA Gotschlich E.C.;
RT "Genetic locus for the biosynthesis of the variable portion of Neisseria gonorrhoeae lipooligosaccharide.";
RL J. Exp. Med. 180:2181-2190(1994).

```

DR EMBL; U14554; AAA68012.1; -.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Transferrase
SQ SEQUENCE 337 AA; 38393 MW; 5AE21AA75CAA732E CRC64;

Query Match 57.5%; Score 1049; DB 2; Length 337;
Best Local Similarity 64.5%; Pred. No. 5,6e-82;

Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

QY 1 MQLPVSLICAYNVEKYFAQSLAAVYVNOTWRNLDILIVDDSGTDLTAIAKDFQKRSRI 60
DB 1 MQLPVSLICAYNVEKYFAQSLAAVYVNOTWRNLDILIVDDSGTDLTAIAKDFQKRSRI 60
QY 61 KILQAOQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIKIYGEEMKDSI 120
DB 61 RIIISPRNLGFIASNLIGDELAKS--GGGEYIARTDADDIASPGWIKIYGEEMKDSI 118
QY 121 IAMGAMLEVLSEKDGNNLRHHKGIWKKPTRHEDIAAEFFPGNPJHNTMIMRSVI 180
DB 119 IAMGAMLEVLSEENKKSVALAIARNGAIWDPTRHEDIVAVFPFGNPJHNTMIMRSVI 178
QY 181 DGLRPTFERMAEDYQWYDVYSKIGRLATYPEALVKRLANOVSSKHSYRQHEIAGI 240
DB 179 DGLRFPDPAVYIHAEDYKFWYAGKIGRLATYPEALVKYRFHODOTSRYNLQQRRTARKI 238
QY 241 OKTARNDELQSMGFETRDSLEYROTAKAAAYELPEKDLPEEDFERARFELYOCFRTDTP 300
DB 239 KEIRIAGYWKAAAGIAGVAGDCLNTGLKSTAYALYKALSGDIGCLRFLYEFYFLSLERY 298
QY 301 PSGAWLDFPADGRMRRLFTLRQYFGILYRLIK 332
DB 299 SLTDLDFLTDRVMKRLFAAPQYRKILKMLR 330

RESULT 13
Q9CLR9 PRELIMINARY; PRT; 337 AA.

AC Q9CLR9;
DT 01-JUN-2001 (TREMblrel. 17, Created)
DT 01-JUN-2001 (TREMblrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Hypothetical protein PM1140.
GN PM1140.
OS Pasteurella multocida.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Pasteurella.
OX NCBI_TaxID=747;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PM70;
RX MEDLINE=21145866; PubMed=11248100;
RA May B.J., Zhang Q., Li L., Paustian M.L., Whittam T.S., Kapur V.;
RT "Complete genomic sequence of Pasteurella multocida Fm70.";
RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 337 AA; 39267 MW; 8639BCFB5F700D84 CRC64;

Query Match 36.6%; Score 666.5; DB 16; Length 337;
Best Local Similarity 42.9%; Pred. No. 4,5e-49;
Matches 145; Conservative 59; Mismatches 113; Indels 21; Gaps 5;

QY 2 QPLVSLICAYNVEKYFAQSLAAVYVNOTWRNLDILIVDDSGTDLTAIAKDFQKRSRI 61
DB 6 QPLVSLICAYNVEKYFAQSLAAVYVNOTWRNLDILIVDDSGTDLTAIAKDFQKRSRI 65
QY 62 ILAQOQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIKIYGEEMKDSI 121
DB 66 IINNEENGFIALSLNIGIASI-----NGDIARTDADDIATPEWIKIYGLYMLSHPOLI 119

QY 122 AMGAMLEVLSEKDGNNLRHHKGIWKKPTRHEDIAAEFFPGNPJHNTMIMRSVI 180
DB 120 AMGAMLEVLSEENKKSVALAIARNGAIWDPTRHEDIVAVFPFGNPJHNTMIMRSVI 179
QY 181 DGLRPTFERMAEDYQWYDVYSKIGRLATYPEALVKRLANOVSSKHSYRQHEIAGI 240
DB 180 EHGLRFPDPAVYIHAEDYKFWYAGKIGRLATYPEALVKYRFHODOTSRYNLQQRRTARKI 239
QY 241 OKTARNDELQSMGFETRDSLEYROTAKAAAYELPEKDLPEEDFERARFELYOCF 294
DB 240 KRRANITLQDLGVYHRLGEDIFPHDITITQAEIASLSLONCI-----IKRLTYCY 292
QY 295 KRTDTPSGAWLDFPADGRMRRLFTLRQYFGILYRLIK 332
DB 293 LSLVQNKILNITLYFLRD--KNSYFENKKQIKIRIIR 329

RESULT 14
P74947 PRELIMINARY; PRT; 337 AA.

AC P74947;
DT 01-FEB-1997 (TREMblrel. 02, Created)
DT 01-MAY-1999 (TREMblrel. 10, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE ORF139-19 protein (Putative glycosyltransferase A).
OS Vibrio cholerae.
OC Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.
OX NCBI_TaxID=666;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MO45;
RA Yamasaki S., Shimizu T., Hoshino K., Ho S., Shimada T., Nair G.B.,
RA Takeda Y.;
RL Submitted (APR-1998) to the EMBL/GenBank/DBJ databases.

RN [2]
RP SEQUENCE OF 157-337 FROM N.A.
RC STRAIN=AI 1838;
RX MEDLINE=97094999; PubMed=8940420;
RA Falkind S., Stark M., Albert M.J., Uhlen M., Lundeberg J.,
RA Weintraub A.;
RT "Cloning and sequence of a region of Vibrio cholerae O139 Bengal and
its use in PCR-based detection.";
RL J. Clin. Microbiol. 34:2904-2908(1996).
DR EMBL; AB012956; BAA33603.1; -.
DR EMBL; U72485; AAC44837.1; -.
DR HSSP; P39621; 10G0.
DR InterPro; IPR001173; Glycos_transf_2.
DR Pfam; PF00535; Glycos_transf_2; 1.
KW Transferrase.
SQ SEQUENCE 337 AA; 39164 MW; 721D748FCB8C9798 CRC64;

Query Match 22.6%; Score 412.5; DB 2; Length 337;
Best Local Similarity 33.0%; Pred. No. 3,2e-27;
Matches 104; Conservative 55; Mismatches 113; Indels 43; Gaps 9;

QY 3 PLVSLICAYNVEKYFAQSLAAVYVNOTWRNLDILIVDDSGTDLTAIAKDFQKRSRI 62
DB 4 PRISVMSVYNGEYKELGALIDSILKQTFSDPEFIINIGSDTKLEIKSYWKKDRI-V 62
QY 63 LAQOQNSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIKIYGEEMKDSI 122
DB 63 LVSRRNKGILVSLNIGDLP-LAK----GQYIARMADDISIKSRFEKQIEFLDSNPDIGV 116
QY 123 MGAMLEVLSEKDGNNLRHHKGIWKKPTRHEDIAAEFFPGNPJHNTMIMRSVIDG 182
DB 117 CGTWVEVGE-----NKSKKMKAPTOPDPLKALFFSVFAHPTWIRKKVYIDK 166
QY 183 GLRPTFERMAEDYQWYDVYSKIGRLATYPEALVKRLANOVSSKHSYRQHEIAGI 237
DB 167 YKINVDLQYKAAEDYKFWVDKSKYTLFNVPEILIRYHQGESISRVADNENKRFELI 226
QY 238 OGIO-----KTARNDELQSMGFETRDSLEYROTAKAAAYELPEKDL 279

Db	227	SKIQNEVLTSIGIYLVTNEGAKNHHIISLNERIINNVYDCMIRAHLLTKISSQIESSQFD	286
Qy	280	EEDFER---ARRFLY	291
		: :	
Db	287	SSAIERLMLKKYFY	301

RESULT	15			
034234				
ID	034234	PRELIMINARY;	PRT;	337 AA.
AC	034234;			
DT	01-JAN-1998	(TREMBLrel. 05, Created)		
DT	01-JAN-1998	(TREMBLrel. 05, Last sequence update)		
DT	01-JUN-2002	(TREMBLrel. 21, Last annotation update)		
DE	Sugar transferase.			
OS	Vibrio cholerae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Vibrionaceae; Vibrio.			
OX	NCBI_TaxID=666;			
RN	[1]			
RP	SEQUENCE FROM N.A.			
RC	STRAIN=AT-1837.			
RX	MEDLINE=97252505; PubMed=9098074;			
RA	Strocher U. H., Parasivam G., Dredge B. K., Manning P. A.;			
RT	"Novel Vibrio cholerae O139 genes involved in lipopolysaccharide			
RT	biosynthesis.";			
RL	J. Bacteriol. 179:2740-2747(1997).			
DR	EMBL: Y07786; CAA69119.1; -.			
DR	HSSP: P39621; 10G0.			
DR	InterPro: IPR001173; Glycos.transf_2.			
DR	Pfam: PF00535; Glycos.transf_2, 1.			
DR	Transferase			
SQ	SEQUENCE 337 AA; 39236 MW; 563DB5188B135464 CRC64;			

Query Match	22.5%	Score 410;	DB 2;	Length 337;
Best Local Similarity	37.9%	Pred. No. 5.3e-27;		
Matches 97;	Conservative 45;	Mismatches 88;	Indels 26;	Gaps 7;

[illegible]

Search completed: December 2, 2002, 12:02:15
Job time : 31.8926 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 31.7609 Seconds

(without alignments)
1283.801 Million cell updates/sec

Title: US-10-007-267-4

Perfect score: 1657

Sequence: 1 MDIVFAADNDNYAAYLCVAAK.....KCMIGWRKKLSARFLRKITY 306

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : A.Geneseq_101002:*

```
1: /SID52/gcgdata/geneseq/geneseq-emb1/AA1980.DAT:*
2: /SID52/gcgdata/geneseq/geneseq-emb1/AA1981.DAT:*
3: /SID52/gcgdata/geneseq/geneseq-emb1/AA1982.DAT:*
4: /SID52/gcgdata/geneseq/geneseq-emb1/AA1983.DAT:*
5: /SID52/gcgdata/geneseq/geneseq-emb1/AA1984.DAT:*
6: /SID52/gcgdata/geneseq/geneseq-emb1/AA1985.DAT:*
7: /SID52/gcgdata/geneseq/geneseq-emb1/AA1986.DAT:*
8: /SID52/gcgdata/geneseq/geneseq-emb1/AA1987.DAT:*
9: /SID52/gcgdata/geneseq/geneseq-emb1/AA1988.DAT:*
10: /SID52/gcgdata/geneseq/geneseq-emb1/AA1989.DAT:*
11: /SID52/gcgdata/geneseq/geneseq-emb1/AA1990.DAT:*
12: /SID52/gcgdata/geneseq/geneseq-emb1/AA1991.DAT:*
13: /SID52/gcgdata/geneseq/geneseq-emb1/AA1992.DAT:*
14: /SID52/gcgdata/geneseq/geneseq-emb1/AA1993.DAT:*
15: /SID52/gcgdata/geneseq/geneseq-emb1/AA1994.DAT:*
16: /SID52/gcgdata/geneseq/geneseq-emb1/AA1995.DAT:*
17: /SID52/gcgdata/geneseq/geneseq-emb1/AA1996.DAT:*
18: /SID52/gcgdata/geneseq/geneseq-emb1/AA1997.DAT:*
19: /SID52/gcgdata/geneseq/geneseq-emb1/AA1998.DAT:*
20: /SID52/gcgdata/geneseq/geneseq-emb1/AA1999.DAT:*
21: /SID52/gcgdata/geneseq/geneseq-emb1/AA2000.DAT:*
22: /SID52/gcgdata/geneseq/geneseq-emb1/AA2001.DAT:*
23: /SID52/gcgdata/geneseq/geneseq-emb1/AA2002.DAT:*
```

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1657	100.0	306	17 AAR91313	N. gonorrhoeae gly
2	1657	100.0	306	18 AAM06578	Lipo-oligosacchari
3	583	35.2	106	22 AAB72456	Partial GlyT. Nei
4	572	34.5	302	21 AAB44538	Virulence gene pro
5	568	34.3	302	21 AAB44538	Human 67210 glycos
6	330	19.9	334	23 AAU9917	Partial GlyT. Hae
7	280	16.9	106	22 AAB72455	Streptococcus pneu
8	233	14.1	399	21 AAY81561	Streptococcus pneu
9	205	12.4	50	20 AAM89327	Neisseria meningit
10	203.5	12.3	401	23 ABB28548	Streptococcus poly

11	197.5	11.9	332	21 AAG15767
12	197.5	11.9	346	21 AAG15766
13	197	11.9	313	21 AAY91285
14	197	11.9	398	23 ABB28550
15	196	11.8	312	23 ABB27411
16	190.5	11.5	332	21 AAG45491
17	190.5	11.5	346	21 AAG45490
18	190.5	11.5	346	23 ABB92808
19	186	11.2	361	23 ABB90816
20	184	11.1	272	23 ABB27661
21	184	11.1	392	21 AAG28027
22	184	11.1	409	21 AAG28026
23	178.5	10.8	441	21 AAG28025
24	176.5	10.7	401	23 ABB27413
25	176.5	10.7	674	19 AAM61238
26	176.5	10.7	674	23 ABB27412
27	175.5	10.6	413	23 ABB27412
28	171.5	10.4	357	21 AAG13328
29	171.5	10.4	357	23 ABB92767
30	171.5	10.4	361	21 AAG13327
31	169.5	10.2	351	23 ABB92218
32	165	10.0	323	21 AAG10403
33	165	10.0	364	21 AAG10402
34	165	10.0	365	21 AAG10401
35	162.5	9.8	341	23 ABB92615
36	157	9.5	305	21 AAG55049
37	157	9.5	361	21 AAG30297
38	157	9.5	372	19 AAM98360
39	157	9.5	372	22 AAB60318
40	157	9.5	384	21 AAG30296
41	157	9.5	390	21 AAG30295
42	157	9.5	390	23 ABB91529
43	156.5	9.4	282	21 AAG13970
44	156.5	9.4	345	21 AAG13969
45	156.5	9.4	365	21 AAG13968

ALIGNMENTS

```
RESULT 1
AAR91313
ID AAR91313 standard; Protein: 306 AA.
AC AAR91313;
DT 09-JUL-1996 (first entry)
XX
DE N. gonorrhoeae glycosyltransferase LgtC.
XX
DE
XX
KW Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;
KW vaccine.
XX
OS Neisseria gonorrhoeae strain F62.
XX
XX WO9610086-A1.
XX
XX
XX 04-APR-1996.
XX
XX 25-SEP-1995; 95WO-US12317.
XX
XX 26-SEP-1994; 94US-0312387.
XX
XX (UTRQ ) UNIV ROCKEFELLER.
XX
XX Gotschlich EC;
XX
XX WPI: 1996-200924/20.
XX
XX N-PSDB; AAT14061.
XX
XX Nucleic acids encoding [glycosyl transferase(s)] - used in the
XX diagnosis of infection with Neisseria and for the biosynthesis of
XX oligo:saccharide(s)
XX
```

Arabidopsis thalia
Arabidopsis thalia
Group B Streptococ
Streptococcus poly
Streptococcus poly
Arabidopsis thalia
Arabidopsis thalia
Herbicidally activ
Streptococcus poly
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Streptococcus poly
S. pneumoniae Sp1
Streptococcus poly
Arabidopsis thalia
Herbicidally activ
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia
Herbicidally activ
Arabidopsis thalia
Arabidopsis thalia
H. pylori GHP0_134
Helicobacter pylori
Arabidopsis thalia
Arabidopsis thalia
Herbicidally activ
Arabidopsis thalia
Arabidopsis thalia
Arabidopsis thalia

XX Claim 10; Fig 2d; 81pp; English.
PS
XX
CC 5 Glycosyltransferases (AA091311-15) are products of the igt locus
CC (AA0914061) of *Neisseria gonorrhoeae* strain F22. Glycosyltransferase
CC 1gtC (AA091313) can be optd. by expression of the 1gtC coding
CC sequence in recombinant host cells. A method for adding gal
CC alpha1-4 to gal comprises contacting a reaction mixture contg.
CC activated gal to an acceptor moiety comprising a gal residue in the
CC presence of 1gtC. Oligosaccharides can be produced that, when
CC attached to non-toxic lipids, are useful for *Neisseria* vaccine prepn.
CC blood group core oligosaccharides, and mimics of lacto-N-neoetrase,
CC gangliosides and saccharide portions of globoglycolipids can also be
CC produced using the enzymes.
XX
SQ Sequence 306 AA;

Query Match 100.0%; Score 1657; DB 17; Length 306;
Best Local Similarity 100.0%; Pred. No. 4.8e-169;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVFAADNDYAAAYLCVAAKSVFAAHPDTEIRPHVLDAGISEENRAVAANLRGGNIRF 60
DB 1 MDIVFAADNDYAAAYLCVAAKSVFAAHPDTEIRPHVLDAGISEENRAVAANLRGGNIRF 60
QY 61 IDVNPEDEGAFPLNIRHISTTTTARLKGLEYIADCKVLYLDVTVRDLKPLMDTDLG 120
DB 61 IDVNPEDEGAFPLNIRHISTTTTARLKGLEYIADCKVLYLDVTVRDLKPLMDTDLG 120
QY 121 GNNVGCACIDFEVROEGYKQKIGMADGEYTFNAGVLLINKKRRHDIKMSCEWEQYK 180
DB 121 GNNVGCACIDFEVROEGYKQKIGMADGEYTFNAGVLLINKKRRHDIKMSCEWEQYK 180
QY 181 DVMOYODDILNGLFKGVCYANSRFPNTNYAFMANGFASRHTDPLYLDRTNTAMPVA 240
DB 181 DVMOYODDILNGLFKGVCYANSRFPNTNYAFMANGFASRHTDPLYLDRTNTAMPVA 240
QY 241 VSHYGSASAKFWHRDCTVWGAEERFTELAGSLTTPVEEMRGLAVPTKCMLOMRKRLSAR 300
DB 241 VSHYGSASAKFWHRDCTVWGAEERFTELAGSLTTPVEEMRGLAVPTKCMLOMRKRLSAR 300
QY 301 FLRKIIY 306
DB 301 FLRKIIY 306

RESULT 2
AAW06578

ID AAW06578 standard; Protein; 306 AA.

AC AAW06578;

DT 21-MAR-1997 (first entry)

DE Lipo-oligosaccharide gene-encoded protein.

KM Polyglycosyltransferase; N-acetylglucosaminyl transferase;

KW N-acetylglactosaminyl transferase; lipo-oligosaccharide.

OS *Neisseria gonorrhoeae* ATCC 33084.

XX WO9640971-A1.

PN 19-DEC-1996.

PF 03-JUN-1996; 96WO-US08323.

PR 07-JUN-1995; 95US-0478140.

PA (NEOS-) NEOSE TECHNOLOGIES INC.

PI Buczala SL, Johnson KF, Roth S;

DR WPI; 1997-052351/05.
DR N-PSDB; AAT49230.

PT Transfer of at least 2 saccharide units using
PT polyglycosyl:transferase - isolated from *N. gonorrhoeae*, catalyses
PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
PT single galactose moiety

PS Disclosure; Fig 2C-E; 38pp; English.

CC A lipo-oligosaccharide-encoding gene region (AAT49230) of *Neisseria*
CC *gonorrhoeae* ATCC 33084 includes coding sequences for 5 proteins
CC (AAW06576-80), one of which (AAW06576) is a polyglycosyltransferase
CC that catalyses the addition of GlcNAc and GalNAc disaccharides to
CC a galactose moiety. The function of the other 4 proteins is not
CC stated in the specification.

SQ Sequence 306 AA;

Query Match 100.0%; Score 1657; DB 18; Length 306;
Best Local Similarity 100.0%; Pred. No. 4.8e-169;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVFAADNDYAAAYLCVAAKSVFAAHPDTEIRPHVLDAGISEENRAVAANLRGGNIRF 60
DB 1 MDIVFAADNDYAAAYLCVAAKSVFAAHPDTEIRPHVLDAGISEENRAVAANLRGGNIRF 60
QY 61 IDVNPEDEGAFPLNIRHISTTTTARLKGLEYIADCKVLYLDVTVRDLKPLMDTDLG 120
DB 61 IDVNPEDEGAFPLNIRHISTTTTARLKGLEYIADCKVLYLDVTVRDLKPLMDTDLG 120
QY 121 GNNVGCACIDFEVROEGYKQKIGMADGEYTFNAGVLLINKKRRHDIKMSCEWEQYK 180
DB 121 GNNVGCACIDFEVROEGYKQKIGMADGEYTFNAGVLLINKKRRHDIKMSCEWEQYK 180
QY 181 DVMOYODDILNGLFKGVCYANSRFPNTNYAFMANGFASRHTDPLYLDRTNTAMPVA 240
DB 181 DVMOYODDILNGLFKGVCYANSRFPNTNYAFMANGFASRHTDPLYLDRTNTAMPVA 240
QY 241 VSHYGSASAKFWHRDCTVWGAEERFTELAGSLTTPVEEMRGLAVPTKCMLOMRKRLSAR 300
DB 241 VSHYGSASAKFWHRDCTVWGAEERFTELAGSLTTPVEEMRGLAVPTKCMLOMRKRLSAR 300
QY 301 FLRKIIY 306
DB 301 FLRKIIY 306

RESULT 3
AAB72456

ID AAB72456 standard; Protein; 106 AA.

AC AAB72456;

DT 08-MAY-2001 (first entry)

DE Partial GLYT.

KW UGCT; UDP-glucose:glycoprotein glucosyltransferase; enzyme; GLYT.

OS *Neisseria gonorrhoeae*.

XX WO200112845-A1.

PN 22-FEB-2001.

PF 27-JUL-2000; 2000WO-CA00883.

PR 18-AUG-1999; 99US-0376330.

PA (CANADA) NAT RES COUNCIL CANADA.

PI Tessier DC, Dignard D, Bergeron JM, Thomas DY;

PT Attenuated Pasteurellaceae bacteria comprising mutations in virulence
PT genes, useful as a live attenuated vaccine against bacterial infections
PT -
XX
PS Disclosure; Pages 149-150; 322pp. English.

PS Disclosure; Pages 149-150; 322pp; English..

CC The family Pasteurellaceae encompasses several pathogens that infect a
CC wide variety of animals. The present invention relates to virulence genes
CC from Pasteurellaceae. The present sequence is a protein encoded by one
CC such virulence gene. The virulence genes of the present invention may be
CC mutated in order to produce an inactive gene. The inactive virulence gene
CC may in turn be used to produce a vaccine, which is useful for treating
CC bacterial infections such as septiciemias, bronchopneumonias, thinitis and
CC wound infections.

SQ Sequence 302 AA;

Query Match	34.3%	Score 568	DB 21	Length 302
Best Local Similarly	38.6%	Pred. No. 3.7e-52		
Matches 120	Conservative 61	Mismatches 116	Indels 14	Gaps 7

QY	1	MDIYFAADDNNAAALCYAAKSVENAAHDDTELREFVLDLAGISEENRAAVANLR--GGNIR	59
Db	1	MNLFVSDSDYAAKHLLVAAKSI--INHNDEKGSIFYFDLGEKDEKRNINDIVSSGSEVN	59
QY	60	FIDVNPDEFGFPLNIRKHSITTTARLKLGEYIND--CDKVLVLDTDVLYVNDGKLPLMDDTD	118
Db	60	FIANEEKEEFSEPOAISTISATYARLKAAYEDLDNKKITLIDVDVLYVNSLEMLNVD	119
QY	119	LGGMWVGACIDLFEVEREG--VKQIGMADGGEYYPNAGVLLINLEKKRRHIDFKMSEWE	177
Db	120	VNNEFLTACQSYSLFENKESEHKHSISMSDKRYYPNAGVYMLENDDEKRMKDDVFSALDLLA	179
QY	178	QYKDVMQYQDDDLINGLFKSGVCYCAANSRFPNPNVYAFMANGFASRTDPLDYLDRTTAM	233
Db	180	MYPMQMTYQDDDLINILFRNKVVCYCLDRFENMPQLEIRKQYHNKXKSSNHSLSLEK--TTM	237
QY	238	PVAASHYCGSAKPMWRPDCTVGARETTELAGSLTTPVEEMRG--KLAVPPTKCLQDPMRK	295
Db	238	PVLSHSHYCGPEKAAHNDCKHNENVLYQKILAXXS-----RGXDKENVLSIKTYLKLALIR	297
QY	296	KLSARFLRKITY 306	
Db	292	RIRYKFKYQYV 302	

RESULT 6
AAU99917
ID AAU99917 standard; Protein; 334 AA.

DT	07-OCT-2002 (first entry)
XX	
DE	Human 67210 glycosyl transferase domain consensus sequence

KW 67210; glucosyl transferase; infection;
KW haematopoietic disorder; blood clotting disorder; cancer;
KW autoimmune disorder; leukaemia; immunological disorder;
KW cardiovascular disorder; neurological disorder; cellular proliferation;
KW red blood cell disorder; viral disease; neurological disorder.

OS Synthetic.

PN W0200240656-A2.

PD 23-MAY-2002.

14-NOV-2001; 2001WO-US45291.

14-NOV-2000; 2000US-248331P.

PR	14-NOV-2000; 2000US-248365P.
PR	14 NOV 2000; 2000US-248365P.

6
3
2

PR	30-NOV-2000;	2000US-250077P
PR	30-NOV-2000;	2000US-250176P
PR	30-NOV-2000;	2000US-250327P
XX		
PA	(MILL-) MILLENNIUM PHARM INC	
XX		
PI	Meyers RE, Curtis RAD, GL	
XX		
DR	WPI; 2002-508325/54.	

PA (MILL-) MILLENNIUM PHARM INC.

PI Meyers RE, Curtis RAJ, Glucksmann MA;

DR WPI; 2002-508325/54.

PT Isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85044
PT or 84234 polypeptides, useful as reagents or targets for treating or
PT diagnosing pain or metabolic, liver, kidney, or cardiovascular
PT disorders -

PS Disclosure; Flg 6; 298pp; English.

CC This invention relates to the DNA and protein sequences of novel
CC Isolated 47476, 67210, 49875, 46842, 33201, 83378, 84233, 64708, 85041
CC or 84234 proteins. The method of the invention is useful for treating a
CC disorder characterised by aberrant activity of 47476, 67210, 49875,
CC 46842, 33201, 83378, 84233, 64708, 85041 or 84234-expressing cell in a
CC subject. The protein molecules can act as novel diagnostic targets and
CC therapeutic agents for controlling aberrant or deficient signal
CC transduction resulting, in e.g., haematopoietic disorders, including
CC blood clotting disorders, autoimmune disorders, or disorders related to
CC an inability to clear infections (e.g., viral or bacterial infections),
CC as well as disorders related to abnormal cellular proliferation or
CC differentiation, e.g., leukaemia. They may also be used to control
CC disorders of metabolic imbalance (e.g., disorders of lipopolysaccharide
CC biosynthesis or glycogen synthesis) immunological disorders,
CC cardiovascular disorders, neurological disorders, or cellular
CC proliferation and/or differentiation disorders, e.g., cancer, cell
CC motility and adhesion disorders differentiative disorders, red blood
CC cell disorders, viral diseases, neurological disorders (e.g., brain
CC disorders), pain or metabolic disorders, liver disorders, kidney
CC disorders, disorders of the small intestine, disorders of metal ion
CC imbalance, protein trafficking disorders and disorders associated with
CC bone metabolism. The sequences of the invention are also useful for
CC screening assays, predictive medicine (e.g., diagnostic assays,
CC prognostic assays, monitoring clinical trials, and pharmacogenetics);
CC and methods of treatment (e.g., therapeutic and prophylactic). The
CC present sequence represents a predicted consensus sequence motif
CC found in the protein of the invention.

...	Sequence	334	AA;
SQ			

Query Match	19.9%	Score 330	DB 23	Length 334
Best Local Similarity	34.9%	Pred. No. 1.5e-26		
Matches 102; Conservative	33	Mismatches 87	Indels 70	Gaps 15

```

QY 28 DTIRFHVLDAGISENNAVAANLRGGN-----IRIDVNPED- F 68
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 17 ENNAVFHILDTGFEKSKSLVLSQ-----GDKNKLLTETLAPPEFQHFIRIDYNTSDELL 72
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 69 AGEPLNTI--RHISITTYAR-----LKIGEYIADCDK- VLYDTDVLVRDG- LKPLMDT 117
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 73 ESLPTANMKRRIRISKATFARIEDIQLLTADLLDPYSSKRRVLYDADVYVKGADISELLDL 132
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 118 DLGGMMV-GAC-----IDLVEYREG----YKQIGMADG-----EYFNAGVL 136
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 133 DLGEEKILIGAEDLVLSNESKEFFFDVQERAKSRILIEEDFLVETGDLKVIQSYFNAGVL 192
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 157 LINTKKRRHDIIFKSCFEWYEQYK- DVMQODDIDJNGLEFG-----GYCYAN 203
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 153 LINTKKREENLFLKALELLNKKKEEDYVLEFPDDVINTLIFKGFKSKDELDISTARVLELP 252
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY 204 SRENFMTPTNYAFMANGASRHTDPLDYLDRTN-TAMPAVASHYCGS- AKPMHR 253
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 253 PRYNTQYNNAVYMYVFLTKRRKEKTLAKKLNPDTGNSIYVYCGAGYTPMK 304
   : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

RESULT 7

AAB72455
ID AAB72455 standard: Protein; 106 AA.
XX
AC AAB72455;
XX
DT 08-MAY-2001 (first entry)
XX
DE Partial GLUT.
XX
KW UGGT; UDP-glucose:glycoprotein glucosyltransferase; enzyme; GLUT.
XX
OS Haemophilus influenzae.
XX
PN MO200112845-A1.
XX
PD 22-FEB-2001.
XX
PF 27-JUL-2000; 2000WO-CA00883.
XX
PR 18-AUG-1999; 99US-0376330.
XX
PA (CANA) NAT RES COUNCIL CANADA.
XX
PI Tessier DC, Dignard D, Bergeron JM, Thomas DY;
XX
DR WPI; 2001-218358/22.
XX
PT Determining the effect of a test sample on UDP-glucose:glycoprotein
XX
PT glucosyltransferase (UGGT), useful for measuring UGGT activity,
XX
PT comprises exposing an acceptor substrate for UGGT to a labeled donor in
XX
PT the presence of UGGT.
XX
PS Disclosure: Fig 10; 95pp; English.
XX
XX The present invention relates to a method for determining the effect of a
XX
XX test sample on UDP-glucose:glycoprotein glucosyltransferase (UGGT)
XX
XX activity. The method comprises exposing an acceptor substrate for UGGT to
XX
XX a labelled donor in the presence of the test sample and UGGT. The method
XX
XX is useful for determining UGGT activity. In particular, the method is
XX
XX useful in glucosyltransferase assay and kinetics measurement for
XX
XX determining UGGT activity. UGGT is a soluble enzyme of the endoplasmic
XX
XX reticulum which catalyses the addition of a glucose residue onto
XX
XX asparagine-linked oligosaccharides, which are present on incorrectly
XX
XX folded glycoproteins. The present sequence was used in a sequence
XX
XX homology comparison with rat UGGT (see AAF60732 and AAB72436) which was
XX
XX used in the method of the present invention.
XX
SQ Sequence 106 AA;
XX
Query Match 16.9%; Score 280; DB 22; Length 106;
Best Local Similarity 46.7%; Pred. No. 6,3e-22;
Matches 50; Conservative 24; Mismatches 31; Indels 2; Gaps 2;
XX
OY 92 IADCDKVLVLDVLDGLKPLMDTDLGNNVACIDLEVE-ROEGYKQKIGMADGEY 150
DB 1 INIKKAIYIDVDTLTNSSLDLMNDITNYKACRDFIVKKNKAYKKTIGL-EGYSY 59
OY 151 ENAGVLLINLKRRHRDIFKMSCWEVEQYKQYODODIILGKRG 197
DB 60 FNAGILLINLWKKEENIFOKSINMNKNYNNVAKYODODILGICKG 106
XX
RESULT 8
AAY81561
ID AAY81561 standard: Protein; 399 AA.
XX
AC AAY81561;
XX
DT 24-MAY-2000 (first entry)
XX
DE Streptococcus pneumoniae type 4 protein sequence #61.
XX
KW Streptococcus pneumoniae; vaccine; screening; protein antigen;

KW antibacterial; antiinflammatory; meningitis; infection; diagnosis;
KW pneumococcal disease.
XX
OS Streptococcus pneumoniae.
XX
PN WO200006737-A2.
XX
PD 10-FEB-2000.
XX
PF 27-JUL-1999; 99WO-GB02451.
XX
PR 27-JUL-1998; 98GB-0016337.
XX
PR 19-MAR-1999; 99US-0125164.
XX
PA (MICR-) MICROBIAL TECHNICS LTD.
XX
PI Gilbert CFG, Hansbro PM;
XX
DR WPI; 2000-195300/17.
XX
PT New Streptococcal protein, useful as a vaccine, for diagnosis of
XX
PT pneumococcal diseases and for screening agents capable of antagonizing
XX
PT or inhibiting expression of the protein
XX
PS Claim 1; Page 78; 108pp; English.
XX
XX AAY81501 to AAY81679 represent specifically claimed protein sequences
XX
XX isolated from Streptococcus pneumoniae. AAY05407 to AAY05590 represent
XX
XX specifically claimed nucleotide sequences isolated from S. pneumoniae.
XX
XX The sequences have antibacterial and antiinflammatory properties.
XX
XX The protein sequences, and fragments of them, are useful as immunogens
XX
XX and/or antigens. The nucleotide sequences can be used in vaccines and in
XX
XX diagnostic assays. The proteins and nucleotides can be useful for the
XX
XX detection and diagnosis of S. pneumoniae. The protein sequences are also
XX
XX useful for screening an agent capable of antagonizing, inhibiting or
XX
XX interfering with the function or expression of the proteins in which the
XX
XX agent is useful for treatment or prophylaxis of S. pneumoniae infection
XX
XX and meningitis. AAY05591 to AAY05614 represent primers used in the
XX
XX exemplification of the present invention.
XX
SQ Sequence 399 AA;
XX
Query Match 14.1%; Score 233; DB 21; Length 399;
Best Local Similarity 27.5%; Pred. No. 5,2e-16;
Matches 74; Conservative 42; Mismatches 115; Indels 38; Gaps 9;
XX
OY 3 IYFAADDNTAAYLCVAAKSVEAHPDTEIRFVLDAGISEE-----NRAVAANLRGGN 57
DB 5 IYLAADNAYLIPLETTIKSV--LYHNRVDVFXILNSDIAPFEMFKLGRKMEYVN---ST 58
OY 58 IRFIDVNPEDFAGFPLNIRHISITTYARKLGEYIADCDKVLVLDVLDVYRGLKPLMDT 117
DB 59 IKSVDHDKLEFESYKIG-PRINYASTFRFPATE-VVESRVLVLDSDIIVTELATLFBI 116
OY 118 DLGMMWVACIDLEVEUROGYKQKIGMADGEYFNAGVLLINLKWRHRDIFKMSCWEVE 177
DB 117 DLKGYSIGAVDVYA--YEGRSKSG-----FNTGMLMDVAKRKKEISYNSLLELAA 165
OY 178 QYKDYMOYODODILNGLEFGKGCYANSRFPNTYAFNANGFASRHDPLYLDRNTAM 237
DB 166 EQNQVYHLDDQSLINLYFEDNMLALDKTYNNVYGIDYIHLAQECERLDD-----N 215
OY 238 PYAVSHYCGSAKPMWRDCTVMGAERFTEL 266
DB 216 PPTIVYASHDKPMW---TYSISRLEL 240
XX
RESULT 9
AAM89327
ID AAM89327 standard: Peptide; 50 AA.
XX
AC AAM89327;
XX

AC AAG15767;
XX
XX 17-OCT-2000 (first entry)
XX
DE Arabidopsis thaliana protein fragment SEQ ID NO: 16147.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
KW hydridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
OS Arabidopsis thaliana.
PN EPI03405-A2.
XX
XX 06-SEP-2000.
PD
XX
XX 25-FEB-2000; 2000EP-0301439.
PF
XX 25-FEB-1999; 99US-0121825.
PR 05-MAR-1999; 99US-0123180.
PR 09-MAR-1999; 99US-0123548.
PR 23-MAR-1999; 99US-0125788.
PR 25-MAR-1999; 99US-0125264.
PR 29-MAR-1999; 99US-0126785.
PR 01-APR-1999; 99US-0127462.
PR 06-APR-1999; 99US-0128234.
PR 08-APR-1999; 99US-0128714.
PR 16-APR-1999; 99US-0129845.
PR 19-APR-1999; 99US-0130077.
PR 21-APR-1999; 99US-0130449.
PR 23-APR-1999; 99US-0130510.
PR 23-APR-1999; 99US-0130891.
PR 28-APR-1999; 99US-0131449.
PR 30-APR-1999; 99US-0132048.
PR 30-APR-1999; 99US-0132407.
PR 04-MAY-1999; 99US-0132484.
PR 05-MAY-1999; 99US-0132485.
PR 06-MAY-1999; 99US-0132486.
PR 06-MAY-1999; 99US-0132487.
PR 07-MAY-1999; 99US-0133863.
PR 11-MAY-1999; 99US-0134256.
PR 14-MAY-1999; 99US-0134218.
PR 14-MAY-1999; 99US-0134219.
PR 14-MAY-1999; 99US-0134221.
PR 14-MAY-1999; 99US-0134370.
PR 18-MAY-1999; 99US-0134768.
PR 19-MAY-1999; 99US-0134941.
PR 20-MAY-1999; 99US-0135124.
PR 21-MAY-1999; 99US-0135353.
PR 24-MAY-1999; 99US-0135629.
PR 25-MAY-1999; 99US-0136021.
PR 27-MAY-1999; 99US-0136392.
PR 28-MAY-1999; 99US-0136782.
PR 01-JUN-1999; 99US-0137222.
PR 03-JUN-1999; 99US-0137528.
PR 04-JUN-1999; 99US-0137502.
PR 07-JUN-1999; 99US-0137724.
PR 08-JUN-1999; 99US-0138094.
PR 10-JUN-1999; 99US-0138540.
PR 10-JUN-1999; 99US-0138647.
PR 14-JUN-1999; 99US-0139119.
PR 16-JUN-1999; 99US-0139452.
PR 16-JUN-1999; 99US-0139453.
PR 17-JUN-1999; 99US-0139492.
PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.

PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 20-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 03-AUG-1999; 99US-0146389.
PR 04-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 05-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 06-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 09-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 10-AUG-1999; 99US-0147935.
PR 11-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.
PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149426.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149925.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.

```
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151438.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152363.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159295.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159337.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160766.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 26-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161992.
PR 28-OCT-1999; 99US-0161993.
PR 28-OCT-1999; 99US-0162142.
PR 29-OCT-1999; 99US-0162142.
```

Query Match 11.9%; Score 197.5; DB 21; Length 332;
Best Local Similarity 26.3%; Pred. No. 2,6e-12;
Matches 76; Conservative 38; Mismatches 104; Indels 71; Gaps 14;

```
OY 3 IVEAADDNAAATYICVAKSV--EAAHPDTEIRPHVLDAGISEENR--AAVAANLRGGNIN 58
DB 54 VAITLDEVLRKSGIAVNSILQHSVCPES-VFFHFL--AVSEETNILESIVRSVFGLKF 110
OY 59 RFLVDNPEDFAGF-----PLNIRHISITTYARLKLGEYIADC-DKVLVLDTDVYL 106
DB 111 NIYDFAPETVRGSISSVQAOLEQPLN-----YARSLADLLEPCVNRVITYLSDLY 162
OY 107 VRDGLPLMDTDIGMVGCA---CIDLEVRQEGYKOKIGMADGEY-----YFNAGVLL 157
DB 163 VYDDIADIKLWKTSLGSRILIGAPERYCANFTKYFTGFGFMSSEERFSGFRGRKPCYFNTGVWV 222
OY 158 INKKRRRHDIFPMSCGEWGEQYKDWQYODODLNL-----FKGGVCANRFRNFMPT 211
DB 223 IDLKKWRGGYTRIKEMMEIQRRERIYE---LGSLLPPLLVFAGHVAIDISHRMN---- 274
```

```
OY 212 NYAFMANGF-----ASRHTDPLYLDRNTAMPVAVSHYGSAAKPMHR 253
DB 275 -----QHIGLDGDNVNRGSCRDLP-----GPVSLHWSGSGKPMWR 309

RESULT 12
AAG15766
ID AAG15766 standard; Protein: 346 AA.
XX
XX AAG15766;
XX
XX 17-OCT-2000 (first entry)
XX
XX Arabidopsis thaliana protein fragment SEQ ID NO: 16146.
XX
XX Protein identification; signal transduction pathway; metabolic pathway;
XX hybridisation assay; genetic mapping; gene expression control; promoter;
XX termination sequence.
XX
XX Arabidopsis thaliana.
XX
XX EP1033405-A2.
XX
XX 06-SEP-2000.
XX
XX 25-FEB-2000; 2000EP-0301439.
XX
XX
XX 25-FEB-1999; 99US-0121825.
XX 05-MAR-1999; 99US-0122180.
XX 09-MAR-1999; 99US-0123548.
XX 23-MAR-1999; 99US-0125788.
XX 23-MAR-1999; 99US-0126264.
XX 29-MAR-1999; 99US-0126785.
XX 01-APR-1999; 99US-0127462.
XX 06-APR-1999; 99US-0128234.
XX 08-APR-1999; 99US-0128714.
XX 16-APR-1999; 99US-0129845.
XX 19-APR-1999; 99US-0130077.
XX 21-APR-1999; 99US-0130449.
XX 23-APR-1999; 99US-0130510.
XX 28-APR-1999; 99US-0130891.
XX 30-APR-1999; 99US-0131449.
XX 30-APR-1999; 99US-0132048.
XX 30-APR-1999; 99US-0132407.
XX 04-MAY-1999; 99US-0132484.
XX 05-MAY-1999; 99US-0132485.
XX 06-MAY-1999; 99US-0132486.
XX 07-MAY-1999; 99US-0132487.
XX 11-MAY-1999; 99US-0132853.
XX 11-MAY-1999; 99US-0134256.
XX 14-MAY-1999; 99US-0134218.
XX 14-MAY-1999; 99US-0134219.
XX 14-MAY-1999; 99US-0134221.
XX 18-MAY-1999; 99US-0134370.
XX 18-MAY-1999; 99US-0134768.
XX 19-MAY-1999; 99US-0134941.
XX 20-MAY-1999; 99US-0135124.
XX 21-MAY-1999; 99US-0135353.
XX 24-MAY-1999; 99US-0135629.
XX 25-MAY-1999; 99US-0136021.
XX 27-MAY-1999; 99US-0136392.
XX 28-MAY-1999; 99US-0136782.
XX 01-JUN-1999; 99US-0137222.
XX 03-JUN-1999; 99US-0137528.
XX 04-JUN-1999; 99US-0137502.
XX 07-JUN-1999; 99US-0137724.
XX 08-JUN-1999; 99US-0138094.
XX 10-JUN-1999; 99US-0138540.
XX 10-JUN-1999; 99US-0138847.
XX 14-JUN-1999; 99US-0139119.
XX 16-JUN-1999; 99US-0139452.
XX 16-JUN-1999; 99US-0139453.
XX 17-JUN-1999; 99US-0139492.
```


PR 18-JUN-1999; 99US-0139454.
PR 18-JUN-1999; 99US-0139455.
PR 18-JUN-1999; 99US-0139456.
PR 18-JUN-1999; 99US-0139457.
PR 18-JUN-1999; 99US-0139458.
PR 18-JUN-1999; 99US-0139459.
PR 18-JUN-1999; 99US-0139460.
PR 18-JUN-1999; 99US-0139461.
PR 18-JUN-1999; 99US-0139462.
PR 18-JUN-1999; 99US-0139463.
PR 18-JUN-1999; 99US-0139750.
PR 18-JUN-1999; 99US-0139763.
PR 21-JUN-1999; 99US-0139817.
PR 22-JUN-1999; 99US-0139899.
PR 23-JUN-1999; 99US-0140353.
PR 23-JUN-1999; 99US-0140354.
PR 24-JUN-1999; 99US-0140695.
PR 28-JUN-1999; 99US-0140823.
PR 29-JUN-1999; 99US-0140991.
PR 30-JUN-1999; 99US-0141287.
PR 01-JUL-1999; 99US-0141842.
PR 01-JUL-1999; 99US-0142154.
PR 02-JUL-1999; 99US-0142055.
PR 06-JUL-1999; 99US-0142390.
PR 08-JUL-1999; 99US-0142803.
PR 09-JUL-1999; 99US-0142920.
PR 12-JUL-1999; 99US-0142977.
PR 13-JUL-1999; 99US-0143542.
PR 14-JUL-1999; 99US-0143624.
PR 15-JUL-1999; 99US-0144005.
PR 16-JUL-1999; 99US-0144085.
PR 16-JUL-1999; 99US-0144086.
PR 19-JUL-1999; 99US-0144325.
PR 19-JUL-1999; 99US-0144331.
PR 19-JUL-1999; 99US-0144332.
PR 19-JUL-1999; 99US-0144333.
PR 19-JUL-1999; 99US-0144334.
PR 19-JUL-1999; 99US-0144335.
PR 20-JUL-1999; 99US-0144352.
PR 20-JUL-1999; 99US-0144632.
PR 21-JUL-1999; 99US-0144884.
PR 21-JUL-1999; 99US-0144814.
PR 21-JUL-1999; 99US-0145086.
PR 21-JUL-1999; 99US-0145088.
PR 22-JUL-1999; 99US-0145085.
PR 22-JUL-1999; 99US-0145087.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145089.
PR 22-JUL-1999; 99US-0145192.
PR 23-JUL-1999; 99US-0145145.
PR 23-JUL-1999; 99US-0145218.
PR 23-JUL-1999; 99US-0145224.
PR 26-JUL-1999; 99US-0145276.
PR 27-JUL-1999; 99US-0145913.
PR 27-JUL-1999; 99US-0145918.
PR 27-JUL-1999; 99US-0145919.
PR 28-JUL-1999; 99US-0145951.
PR 02-AUG-1999; 99US-0146386.
PR 02-AUG-1999; 99US-0146388.
PR 02-AUG-1999; 99US-0146389.
PR 03-AUG-1999; 99US-0147038.
PR 04-AUG-1999; 99US-0147204.
PR 04-AUG-1999; 99US-0147302.
PR 05-AUG-1999; 99US-0147192.
PR 05-AUG-1999; 99US-0147260.
PR 06-AUG-1999; 99US-0147303.
PR 06-AUG-1999; 99US-0147416.
PR 09-AUG-1999; 99US-0147493.
PR 09-AUG-1999; 99US-0147935.
PR 10-AUG-1999; 99US-0148171.
PR 11-AUG-1999; 99US-0148319.
PR 12-AUG-1999; 99US-0148341.
PR 13-AUG-1999; 99US-0148565.
PR 13-AUG-1999; 99US-0148684.

PR 16-AUG-1999; 99US-0149368.
PR 17-AUG-1999; 99US-0149175.
PR 18-AUG-1999; 99US-0149175.
PR 20-AUG-1999; 99US-0149722.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149723.
PR 20-AUG-1999; 99US-0149729.
PR 23-AUG-1999; 99US-0149902.
PR 23-AUG-1999; 99US-0149930.
PR 25-AUG-1999; 99US-0150566.
PR 26-AUG-1999; 99US-0150884.
PR 27-AUG-1999; 99US-0151065.
PR 27-AUG-1999; 99US-0151066.
PR 27-AUG-1999; 99US-0151080.
PR 30-AUG-1999; 99US-0151303.
PR 31-AUG-1999; 99US-0151303.
PR 01-SEP-1999; 99US-0151930.
PR 07-SEP-1999; 99US-0152263.
PR 10-SEP-1999; 99US-0153070.
PR 13-SEP-1999; 99US-0153758.
PR 15-SEP-1999; 99US-0154018.
PR 16-SEP-1999; 99US-0154039.
PR 20-SEP-1999; 99US-0154779.
PR 22-SEP-1999; 99US-0155139.
PR 23-SEP-1999; 99US-0155486.
PR 24-SEP-1999; 99US-0155659.
PR 28-SEP-1999; 99US-0156458.
PR 29-SEP-1999; 99US-0156596.
PR 04-OCT-1999; 99US-0157117.
PR 05-OCT-1999; 99US-0157753.
PR 06-OCT-1999; 99US-0157865.
PR 07-OCT-1999; 99US-0158029.
PR 08-OCT-1999; 99US-0158232.
PR 12-OCT-1999; 99US-0158369.
PR 13-OCT-1999; 99US-0158293.
PR 13-OCT-1999; 99US-0159294.
PR 13-OCT-1999; 99US-0159299.
PR 14-OCT-1999; 99US-0159329.
PR 14-OCT-1999; 99US-0159330.
PR 14-OCT-1999; 99US-0159331.
PR 14-OCT-1999; 99US-0159637.
PR 14-OCT-1999; 99US-0159638.
PR 18-OCT-1999; 99US-0159584.
PR 21-OCT-1999; 99US-0160741.
PR 21-OCT-1999; 99US-0160767.
PR 21-OCT-1999; 99US-0160768.
PR 21-OCT-1999; 99US-0160770.
PR 21-OCT-1999; 99US-0160814.
PR 21-OCT-1999; 99US-0160815.
PR 22-OCT-1999; 99US-0160980.
PR 22-OCT-1999; 99US-0160981.
PR 22-OCT-1999; 99US-0160989.
PR 25-OCT-1999; 99US-0161404.
PR 25-OCT-1999; 99US-0161405.
PR 25-OCT-1999; 99US-0161406.
PR 26-OCT-1999; 99US-0161359.
PR 26-OCT-1999; 99US-0161360.
PR 26-OCT-1999; 99US-0161361.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161920.
PR 28-OCT-1999; 99US-0161993.
PR 29-OCT-1999; 99US-0162142.

Query Match 11.9%; Score 197.5; DB 21; Length 346;
Best Local Similarity 26.3%; Pred. No. 2,7e-12;
Matches 76; Conservative 38; Mismatches 104; Indels 71; Gaps 14;

QY 3 IVEFADNVAAYLCVAKSV--EAAHPDEIRFHVLDAGISSENR--AAVANAIRGGNI 58
DB 68 VATTLDVEYLRGSIANVSLQHSVCPES-VFFHFI--AVSEETNLESIVRSVFPGLKF 124
QY 59 RFTDVNPEDFAGF-----PLNIRHISTTTTARLKLGEYIADC-DKVLXLTDTVL 106
DB 125 NIYDFAPETVRGLISSVROALEOPLN-----YARSYLADLLPCVNRVYILDSDLV 176

QY 58 IRFIDVNPEDFAGFPLNIRHISITTYARLKLGEYIADCCKVLYLDTDLVLRDGLKPLMDT 117

used in gene therapy: antibodies cc (1) are used for antigen chromatography, immunoassays, and distinguishing/identifying

CC Streptococcus proteins.

```
XX Sequence 398 AA;
SQ
Query Match 11.9%; Score 197; DB 23; Length 398;
Best Local Similarity 28.1%; Pred. No. 3.8e-12;
Matches 73; Conservative 42; Mismatches 103; Indels 42; Gaps 11;
OY 3 IVFAADNYAAVLCVAAKSVENAHPTREIRFHVLDAGISEENRAVAANLKGGN----- 57
DB 7 VVLGADSYIINQIETTLKSLCVYHENLSI--FIFMODIPQEMFLAMKRVQGTGQIDV 64
OY 58 IRFDVNPEDFAGPPLNIRHISITTYARLKGEYADCDKVLVLDPTDVLVRDGLKPLMDT 117
DB 65 KLFHDHLSPKWENKTLN--HINMYTARYETIPQYIS-ADTVLALYDSDLDLVTTNLDNLFQI 121
OY 118 DLGGMVAGACIDLFEVEREGYKOKIKGMADGEYFPMAGVLLINLKKRRHDIFFKSCWEVE 177
DB 122 SLDNAYTLAAVPALE-----GLG---YFNAGVAVINNOQRQEN--MTIKLIE 164
OY 178 QYKDVMOYQ---DQDILNGLFKGGVCYANSRPNMPTNYAFMANGF---ASRHTDPLXLD 231
DB 165 KNQKRIEVAANGSDQTLNRMFENQYIYLDYRNQ-----IGFDGMAIDGKHKFTFD 216
OY 232 RRTNAMPYAVASHYCGSAKPW 251
DB 217 IPITPLP-KIHYYISGIKPW 235

RESULT 15
ABP27411
ID ABP27411 standard; Protein; 312 AA.
XX
AC ABP27411;
XX
DT 02-JUL-2002 (first entry)
XX
DE Streptococcus polypeptide SEQ ID NO 3998.
XX
KW Streptococcus; GAS; GBS; group B streptococcus; Streptococcus agalactiae;
KW group A streptococcus; Streptococcus pyogenes; antibacterial;
KW antiinflammatory; infection; vaccine; meningitis; gene therapy.
XX
OS Streptococcus agalactiae.
XX
PN MO20234771-A2.
XX
PD 02-MAY-2002.
XX
PF 29-OCT-2001; 2001WO-GB04789.
XX
PR 27-OCT-2000; 2000GB-0026333.
PR 24-NOV-2000; 2000GB-0028727.
PR 07-MAR-2001; 2001GB-0005640.
XX
PA (CHIR-) CHIRON SPA.
PA (GENO-) INST GENOMIC RES.
XX
PI Telford J, Masignani V, Margarit Ros YI, Grandi G, Fraser C;
PI Tettelin H;
XX
DR WPI: 2002-352536/38.
DR N-PSDB; ABN68042.
XX
PT New Streptococcus protein for the treatment or prevention of infection
PT or disease caused by Streptococcus bacteria, such as meningitis, and
PT for detecting a compound that binds to the protein -
XX
PS Claim 1; Page 3555; 4525pp; English.
XX
CC The invention relates to a protein (ABP25413-ABP30895) from group B
CC streptococcus/GBS (Streptococcus agalactiae) or group A streptococcus/GAS
CC (Streptococcus pyogenes), comprising one of 5483 sequences (S1), given in
CC the specification. The proteins have antibacterial and antiinflammatory
```

```
CC activity. (I), nucleic acids encoding (I), ABN66044-ABN71526 and
CC antibodies that bind (I) are used in the manufacture of medicaments for
CC the treatment or prevention of infection or disease caused by
CC Streptococcus bacteria, particularly S. agalactiae and S. pyogenes.
CC Nucleic acids encoding (I) are used to detect Streptococcus in a
CC biological sample. (I) is used to determine whether a compound binds to
CC (I). A composition comprising (I) or a nucleic acid encoding (I), may be
CC used as a vaccine or diagnostic composition. The disease caused by
CC Streptococcus that is prevented or treated may be meningitis. Nucleic
CC acid encoding (I) may be used to recombinantly produce (I) and may be
CC used in gene therapy. Antibodies to (I) are used for affinity
CC chromatography, immunoassays, and distinguishing/identifying
CC Streptococcus proteins.
XX
SQ Sequence 312 AA;
Query Match 11.8%; Score 196; DB 23; Length 312;
Best Local Similarity 33.0%; Pred. No. 3.4e-12;
Matches 61; Conservative 22; Mismatches 70; Indels 32; Gaps 8;
OY 83 YARLKGEYIADCDKVLVLDPTDVLVRDGLKPLMDIDLGNNVAGACIDLFVERQEGYKQKI 142
DB 3 YARTYIPQLT-DAEKVLYLDDITLVVDNLDKLFETELDGYPLAATLD----- 48
OY 143 GMADEGEYFPMAGVLLINLKKRRHDIFFKSCWEVBOYKDVMOYODILNGLFKGGCYA 202
DB 49 --GDG-IYFNSGVMLINSLWMRYRVEKLEITERELDNGIFGDQVNLFLFDNNMLKL 105
OY 203 NSRFNEMPTN--YAFMANGFASRHTDPLXLDRTNAMPYAVASHYCGSAKPMHRDCTVGA 260
DB 106 EDKYNAQVGNLGAPEYEWQG-----YFDR-NSEPTII-HYCTHDKPMN---TFSS 152
OY 261 ERFTE 265
DB 153 SRFRE 157
```

Search completed: December 2, 2002, 11:59:01
Job time : 33.7609 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 11.3901 Seconds
(without alignments)
790.458 Million cell updates/sec

Title: US-10-007-267-4

Perfect score: 1657
Sequence: 1 MDIVFAADNDYAAVLCVAAK.....KCMIDRMKKLSARFLRKIV 306

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_AA.*

1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCURS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfilest1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1657	100.0	306	1 US-08-312-387B-4	Sequence 4, Appli
2	1657	100.0	306	1 US-08-683-426-4	Sequence 4, Appli
3	1657	100.0	306	1 US-08-683-458-4	Sequence 4, Appli
4	1657	100.0	306	2 US-08-878-360-4	Sequence 4, Appli
5	1657	100.0	306	3 US-08-478-140B-4	Sequence 4, Appli
6	1657	100.0	306	4 US-09-333-412-4	Sequence 4, Appli
7	1657	100.0	306	4 US-09-338-943-4	Sequence 4, Appli
8	583	35.2	106	4 US-09-376-330-28	Sequence 28, Appli
9	280	16.9	106	4 US-09-376-330-27	Sequence 27, Appli
10	176.5	10.7	674	4 US-08-961-083-200	Sequence 200, App
11	140.5	8.5	105	4 US-09-376-330-29	Sequence 29, Appli
12	132	8.0	328	4 US-09-073-297-6	Sequence 6, Appli
13	130	7.8	106	4 US-09-376-330-25	Sequence 25, Appli
14	121.5	7.3	121	4 US-09-376-330-30	Sequence 30, Appli
15	115	6.9	110	4 US-09-376-330-24	Sequence 24, Appli
16	110.5	6.7	107	4 US-09-376-330-26	Sequence 26, Appli
17	110.5	6.7	1548	4 US-09-376-330-15	Sequence 15, Appli
18	109	6.6	110	4 US-09-376-330-23	Sequence 23, Appli
19	98	5.9	346	4 US-09-073-297-2	Sequence 2, Appli
20	91.5	5.5	115	4 US-09-376-330-20	Sequence 20, Appli
21	89.5	5.4	1327	4 US-09-376-330-2	Sequence 2, Appli
22	87	5.3	394	4 US-09-055-765-2	Sequence 2, Appli
23	87	5.3	576	1 US-08-221-817-22	Sequence 22, Appli
24	87	5.3	576	1 US-08-454-439-22	Sequence 22, Appli
25	87	5.3	576	5 PCR-US94-10487-22	Sequence 22, Appli
26	86	5.2	1447	4 US-09-376-330-17	Sequence 17, Appli
27	81	4.9	389	4 US-09-055-765-15	Sequence 15, Appli

28	80.5	4.9	437	4 US-08-961-083-70	Sequence 70, Appli
29	78.5	4.7	1019	4 US-09-434-066-23	Sequence 23, Appli
30	78	4.7	566	4 US-09-073-297-5	Sequence 5, Appli
31	78	4.7	1041	1 US-08-220-151-4	Sequence 4, Appli
32	78	4.7	1041	1 US-08-413-118-4	Sequence 4, Appli
33	78	4.7	1041	3 US-08-473-446-4	Sequence 4, Appli
34	78	4.7	1702	4 US-08-296-791-5	Sequence 5, Appli
35	78	4.7	1702	5 PCR-US95-10661A-5	Sequence 5, Appli
36	78	4.7	2802	4 US-09-542-331-1	Sequence 1, Appli
37	78	4.7	2802	4 US-09-510-791-1	Sequence 1, Appli
38	77	4.6	1493	4 US-09-376-330-16	Sequence 16, Appli
39	77	4.6	1581	4 US-08-726-320-3	Sequence 3, Appli
40	77	4.6	1581	4 US-09-208-716-3	Sequence 3, Appli
41	77	4.6	1835	3 US-08-836-325-15	Sequence 15, Appli
42	76.5	4.6	117	4 US-09-376-330-34	Sequence 34, Appli
43	76.5	4.6	1066	2 US-08-633-770A-1	Sequence 1, Appli
44	76	4.6	754	4 US-09-521-780-2	Sequence 2, Appli
45	76	4.6	754	4 US-09-521-780-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-08-312-387B-4
; Sequence 4, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESS: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5600
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-312-387B-4

Query Match      100.0%; Score 1657; DB 1; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.9e-171;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

QY 1 MDIVFAADNDYAAVLCVAAKVEAAHPDTEIRPHVLDAGISEENRAVAANRGCGNIRF 60
DB 1 MDIVFAADNDYAAVLCVAAKVEAAHPDTEIRPHVLDAGISEENRAVAANRGCGNIRF 60
QY 61 IDVNPEDFAGPPLNIRHISITTYARLKGELYADCDKVLVLDTVLVRDGLKPLMDTDLG 120

QY	61	IDVNEPFAGPLNIRHISTTTYARLKGELYADCKDVLDPDHDVLRBGLAKLMPDGL	120
Db	61	IDVNEPFAGPLNIRHISTTTYARLKGELYADCKDVLDPDHDVLRBGLAKLMPDGL	120
QY	121	GNWVACIDLFEVERESYKOKIGMADEYYFNAGVLLINLKKWRHDIFFMSCEWWEQYK	180
Db	121	GNWVACIDLFEVERESYKOKIGMADEYYFNAGVLLINLKKWRHDIFFMSCEWWEQYK	180
QY	181	DVMOYODIDLNGLFKSGVYANSRENEMPTNAFMANGEASRHTDPLYIDRNTTAMPVA	240
Db	181	DVMOYODIDLNGLFKSGVYANSRENEMPTNAFMANGEASRHTDPLYIDRNTTAMPVA	240
QY	241	VSHHSGSAKPMHRCSTYWGAEERFELAGSLTTYPEERKGLAAPPRTCMLORRRKTISAR	300
Db	241	VSHHSGSAKPMHRCSTYWGAEERFELAGSLTTYPEERKGLAAPPRTCMLORRRKTISAR	300
QY	301	FLRRKI 306	
Db	301	FLRRKI 306	

QY 61 IDVNEDEAGFPLNIRHISITTYARLKGEYIADCDKYLDTDTVLVBDGLKPLMDTDLG 120
| | | | |
Db 61 IDVNEDEAGFPLNIRHISITTYARLKGEYIADCDKYLDTDTVLVBDGLKPLMDTDLG 120
QY 121 GNMVACIDLEFVEROEGYKOKIGMADGEYFNAGVLLINLKKRRHDFKMSCENEVOYK 180
| | | | |
Db 121 GNMVACIDLEFVEROEGYKOKIGMADGEYFNAGVLLINLKKRRHDFKMSCENEVOYK 180
QY 181 DVMQYODDILNGLEFKGGVCYANSRNFNPTNYAFMANGFASRHTDPLYLDRNTAMPYA 240
| | | | |
Db 181 DVMQYODDILNGLEFKGGVCYANSRNFNPTNYAFMANGFASRHTDPLYLDRNTAMPYA 240
QY 241 VSHYGSAPMHRDCTVWGAERFTELAGSLTTPPEMRGKLAVPTKCLQWRKKLSAR 300
| | | | |
Db 241 VSHYGSAPMHRDCTVWGAERFTELAGSLTTPPEMRGKLAVPTKCLQWRKKLSAR 300
QY 301 FLRKTY 306
| | | | |
Db 301 FLRKTY 306

RESULT 4

US-08-878-360-4
; Sequence 4, Application US/08878360
; Patent No. 5945322
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,360
; FILING DATE: 18-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/683,426
; FILING DATE:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-878-360-4

Query Match 100.0%; Score 1657; DB 2; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.9e-171;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVEADDNTAAVLCVAAKSVAAHPDTEIRFHVLDAGISEENRAVAANIRGGGNIRF 60
| | | | |
Db 1 MDIVEADDNTAAVLCVAAKSVAAHPDTEIRFHVLDAGISEENRAVAANIRGGGNIRF 60
QY 61 IDVNEDEAGFPLNIRHISITTYARLKGEYIADCDKYLDTDTVLVBDGLKPLMDTDLG 120
| | | | |
Db 61 IDVNEDEAGFPLNIRHISITTYARLKGEYIADCDKYLDTDTVLVBDGLKPLMDTDLG 120
QY 121 GNMVACIDLEFVEROEGYKOKIGMADGEYFNAGVLLINLKKRRHDFKMSCENEVOYK 180
| | | | |
Db 121 GNMVACIDLEFVEROEGYKOKIGMADGEYFNAGVLLINLKKRRHDFKMSCENEVOYK 180
QY 181 DVMQYODDILNGLEFKGGVCYANSRNFNPTNYAFMANGFASRHTDPLYLDRNTAMPYA 240
| | | | |
Db 181 DVMQYODDILNGLEFKGGVCYANSRNFNPTNYAFMANGFASRHTDPLYLDRNTAMPYA 240
QY 241 VSHYGSAPMHRDCTVWGAERFTELAGSLTTPPEMRGKLAVPTKCLQWRKKLSAR 300
| | | | |
Db 241 VSHYGSAPMHRDCTVWGAERFTELAGSLTTPPEMRGKLAVPTKCLQWRKKLSAR 300
QY 301 FLRKTY 306
| | | | |
Db 301 FLRKTY 306

RESULT 5

US-08-478-140B-4
; Sequence 4, Application US/08478140B
; Patent No. 6127153
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, KARL F.
; APPLICANT: ROTH, STEPHEN
; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
; TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESSES:
; ADDRESS: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: PC-DOS/MS-DOS
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478,140B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 4:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 306 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-478-140B-4

Query Match 100.0%; Score 1657; DB 3; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.9e-171;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

US-09-338-943-4

Query Match 100.0%; Score 1657; DB 4; Length 306;
Best Local Similarity 100.0%; Pred. No. 9.9e-171;
Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MDIVEAADNNAAYLCVAKSVEAAHPTIEIRFHVLDAGISEENRAAVALRGGINIF 60
DB 1 MDIVEAADNNAAYLCVAKSVEAAHPTIEIRFHVLDAGISEENRAAVALRGGINIF 60
OY 61 IDVNEDDAGFPLNRHSITTYARLKGEYIADCDKVLVLDTVLVKGLPLMDTDLG 120
DB 61 IDVNEDDAGFPLNRHSITTYARLKGEYIADCDKVLVLDTVLVKGLPLMDTDLG 120
OY 121 GNMVACIDLFVEROEGYKOKIGMADGEYFNAGVLLINLKKRRHDIKMSCEWVEQYK 180
DB 121 GNMVACIDLFVEROEGYKOKIGMADGEYFNAGVLLINLKKRRHDIKMSCEWVEQYK 180
OY 181 DVMQYODDILNGLFKGVCYANSRPNEMPTNYAFMANGFASRHTDPLYLDRTNTAMPYA 240
DB 181 DVMQYODDILNGLFKGVCYANSRPNEMPTNYAFMANGFASRHTDPLYLDRTNTAMPYA 240
OY 241 VSHYGSAPMHRDCTVWGAEERETELAGSLTTPVEPMRGKLAVPPTKMLQWRKKLSAR 300
DB 241 VSHYGSAPMHRDCTVWGAEERETELAGSLTTPVEPMRGKLAVPPTKMLQWRKKLSAR 300
OY 301 FLRKTY 306
DB 301 FLRKTY 306

RESULT 8

US-09-376-330-28
; Sequence 28, Application US/09376330
; Patent No. 6399321

; GENERAL INFORMATION:
; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dignard, Daniel
; APPLICANT: Bergeron, John J.M.
; APPLICANT: Thomas, David Y.
; TITLE OF INVENTION: Method for screening for
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity
; FILE REFERENCE: 2139-9"US"
; CURRENT FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 28
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: GLYT-N, gon
US-09-376-330-28

Query Match 35.2%; Score 583; DB 4; Length 106;
Best Local Similarity 100.0%; Pred. No. 2.1e-55;
Matches 106; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 92 IADCKVLYLDTVLVLDGLKPLMDTDLGNNVGACIDLFVEROEGYKOKIGMADGEYF 151
DB 1 IADCKVLYLDTVLVLDGLKPLMDTDLGNNVGACIDLFVEROEGYKOKIGMADGEYF 60
OY 152 NAGVLLINLKKRRHDIKMSCEWVEQYKDVQYODDILNGLFKG 197
DB 61 NAGVLLINLKKRRHDIKMSCEWVEQYKDVQYODDILNGLFKG 106

RESULT 9

US-09-376-330-27
; Sequence 27, Application US/09376330
; Patent No. 6399321

; GENERAL INFORMATION:
; APPLICANT: Tessier, Daniel C.
; APPLICANT: Dignard, Daniel
; APPLICANT: Bergeron, John J.M.
; APPLICANT: Thomas, David Y.
; TITLE OF INVENTION: Method for screening for
; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity
; FILE REFERENCE: 2139-9"US"
; CURRENT FILING DATE: 1999-08-18
; NUMBER OF SEQ ID NOS: 35
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 27
; LENGTH: 106
; TYPE: PRT
; ORGANISM: Artificial Sequence
; FEATURE:
; OTHER INFORMATION: GLUT-H, Inf
US-09-376-330-27

Query Match 16.9%; Score 280; DB 4; Length 106;
Best Local Similarity 46.7%; Pred. No. 1.1e-22;
Matches 50; Conservative 24; Mismatches 31; Indels 2; Gaps 2;

OY 92 IADCKVLYLDTVLVLDGLKPLMDTDLGNNVGACIDLFV-ROEGYKOKIGMADGEY 150
DB 1 IKNIKAIYIDVDTLNSLOGLMNDITNYLACROFIDYKNAKYKTKTGL-EGY 59
OY 151 ENAGVLLINLKKRRHDIKMSCEWVEQYKDVQYODDILNGLFKG 197
DB 60 ENAGVLLINLKKRRHDIKMSCEWVEQYKDVQYODDILNGLFKG 106

RESULT 10

US-08-961-083-200
; Sequence 200, Application US/08961083
; Patent No. 6159469

; GENERAL INFORMATION:
; APPLICANT: Choi et. al.
; TITLE OF INVENTION: Streptococcus pneumoniae Antigen and Vaccines
; NUMBER OF SEQUENCES: 452
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/961,083
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Brookes, A. Anders
; REGISTRATION NUMBER: PB340P2
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 200:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 674 amino acids
; TYPE: amino acid
; STRANDEDNESS: single

TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-200

Query Match 10.7%; Score 176.5; DB 4; Length 674;
Best Local Similarity 25.1%; Pred. No. 2.6e-10;

Matches 72; Conservative 43; Mismatches 107; Indels 65; Gaps 11;

QY 3 IYFAADNTAAALCVAAKSEVAHPDTEIRFHYLDAGISEENRAVAANLRGGNIRFD 62
DB 283 IYLAANYGVDOVLTITKSI--CYHNRSIRFYLIHSDFRNE-----WKQLN 327
QY 63 VNEDFAGPPLNIR-----HISTTYARLKGEYIADCDVLYLDVLRDGL 111
DB 328 KRLKEPDSIELNCRVTSQISCYKSDISYVFLRYFIADVQGE--DKALYDCCDLVVTKNL 386
QY 112 KRLMDPDLGGNMGACIDLFVEROEGYKOKIGMA--DGEYFNAGVLLINKKRRRDIK 170
DB 387 DDLEFATDLDYPLAAVRDGC-----GRAVFGQELFNAGVLLVNNAPFKKKENMTQ 435
QY 171 MSCWEYQYKDWQOXODDILNGLFKGVCYANSRFENMPTNYAFMANGFASHTDPLYL 230
DB 436 KLIDVYENHDKQADQSLNMLFEHKMLELDFOYNHVIYHKQFADYQLPEGQDYP--- 492
QY 231 DRTNMPVAVSHYCGSAKPW-----HRDCTVW---GAERFTEL 266
DB 493 -----AIHYLSHRKPMKDLAQTYYRE--VMWYHGLE-WTEL 527

RESULT 11

US-09-376-330-29
Sequence 29, Application US/09376330
Patent No. 6399321
GENERAL INFORMATION:
APPLICANT: Tessier, Daniel C.
APPLICANT: Dignard, Daniel C.
APPLICANT: Bergeron, John J.M.
TITLE OF INVENTION: Method for screening for
TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity
TITLE OF INVENTION: and nucleic acid encoding for UGT
FILE REFERENCE: 2139-9YUS
CURRENT APPLICATION NUMBER: US/09/376,330
CURRENT FILING DATE: 1999-08-18
NUMBER OF SEQ ID NOS: 35
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 29
LENGTH: 105
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: GSPA-B, sub
US-09-376-330-29

Query Match 8.5%; Score 140.5; DB 4; Length 105;
Best Local Similarity 28.0%; Pred. No. 1.3e-07;

Matches 28; Conservative 27; Mismatches 40; Indels 5; Gaps 2;

QY 97 KVLVLTDLVLRDGLKPLMDTDLGGMWVACIDLFVEROEGYKOKIGMADGEYFNAGVYL 156
DB 6 RMYLIDCDALVLEDISKMDLDIAFTVAAYED---AGQHERLKEKNAVDTDTGTFNSGIM 62
QY 157 LNLKMKRRRHDIFKMSCWEVEQY--KDVMOYODODILNGL 194
DB 63 IIDEFSMRKONITEKYINFINEHPDEDPLVLDQDALNMI 102

RESULT 12

US-09-073-297-6
Sequence 6, Application US/09073297
Patent No. 6255114
GENERAL INFORMATION:
APPLICANT: LIGHTNER, JONATHAN EDWARD

TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898

COMPUTER READABLE FORM:

MEDIUM TYPE: DISKETTE, 3.50 INCH

COMPUTER: IBM PC COMPATIBLE

OPERATING SYSTEM: MICROSOFT WINDOWS 95

SOFTWARE: MICROSOFT WORD VERSION 7.0A

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/09/073,297

FILING DATE:

CLASSIFICATION:

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/852615

FILING DATE: MAY 7, 1997

ATTORNEY/AGENT INFORMATION:

NAME: MAJARIAN, WILLIAM R.

REGISTRATION NUMBER: 41,173

REFERENCE/DOCKET NUMBER: BB-1083-B

TELECOMMUNICATION INFORMATION:

TELEPHONE: 302-992-4926

TELEFAX: 302-773-0164

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 328 amino acids

TYPE: amino acid

STRANDEDNESS: not relevant

TOPOLOGY: linear

MOLECULE TYPE: peptide

US-09-073-297-6

Query Match 8.0%; Score 132; DB 4; Length 328;
Best Local Similarity 23.1%; Pred. No. 5.6e-06;

Matches 56; Conservative 42; Mismatches 64; Indels 80; Gaps 15;

QY 80 ITTYARLKGEYIADCDVLYLDVLRDGLKPLMDTDLGGMWVACIDLFVER----- 134
DB 95 VINYSKRLIWEFY-EXERYVYLDADIOVPDNDIDFLDKGAFY--AVKDCCEKTSHT 151
QY 135 ---QEGYKO---KIGMAGE-----YFNAGVLLINKKRRRHDIFKMSCWEVQYKD 181
DB 152 POTDICTCOQRPDEVAMPRELGPPLTYFNAGMFV-----HE-----PGISTAKD 197
QY 182 VMQ-----YQDODILNGLFKGVCYANSRFENMPTNYAFMANGFASHTDPLYLDR 232
DB 198 LLDALVVTPTFPFAEDPLNMFER-----EQYKIPVYVNLVL-AMLRRHPENVDLDO 249
QY 233 TTNAMPVAVSHCGS-AKQWHRDCTYWGAEFTELAGSLTYPEERKGLAVPTKCMIQ 291
DB 250 -----VKVHYCAAGSKPW-----RFT---GKEENNRE-----DIKMLVK 282
QY 292 RW 293
DB 283 RW 284

RESULT 13

US-09-376-330-25
Sequence 25, Application US/09376330
Patent No. 6399321

GENERAL INFORMATION:

APPLICANT: Tessier, Daniel C.

APPLICANT: Dignard, Daniel

APPLICANT: Bergeron, John J.M.

APPLICANT: Thomas, David Y.

TITLE OF INVENTION: Method for screening for

TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGGT) activity

;; TITLE OF INVENTION: and nucleic acid encoding for UGCT
;; FILE REFERENCE: 2139-9"US"
;; CURRENT APPLICATION NUMBER: US/09/376,330
;; CURRENT FILING DATE: 1999-08-18
;; NUMBER OF SEQ ID NOS: 35
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 25
;; LENGTH: 106
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: GLUT-E. col
US-09-376-330-25

Query Match 7.8%; Score 130; DB 4; Length 106;
Best Local Similarity 33.3%; Pred. No. 1.8e-06;
Matches 35; Conservative 20; Mismatches 44; Indels 6; Gaps 3;

QY 96 DKVLYDPTDVLVRDGLKPLMDTDLGNNVVGACIDLFVEREGYKOKIGMADGE---YYFN 152
DB 5 DRLVLDADVYCKGDISQLHLGLNGA-VAAVVKDVEPMQE--KAVSRLSDELLGQYFN 61
QY 153 AGVLLINKKRRRHIFKMSCWEQYKDVMOYQODDILNGLFKG 197
DB 62 SCVVYLDLKKMADAKLTERKALSLMSKDVYKYPDQDVANNVLLKG 106

RESULT 14

US-09-376-330-30
;; Sequence 30, Application US/09376330
;; Patent No. 6399321
;; GENERAL INFORMATION:
;; APPLICANT: Tessier, Daniel C.
;; APPLICANT: Dignard, Daniel
;; APPLICANT: Bergeron, John J.M.
;; APPLICANT: Thomas, David Y.
;; TITLE OF INVENTION: Method for screening for
;; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGCT) activity
;; TITLE OF INVENTION: and nucleic acid encoding for UGCT
;; FILE REFERENCE: 2139-9"US"
;; CURRENT APPLICATION NUMBER: US/09/376,330
;; CURRENT FILING DATE: 1999-08-18
;; NUMBER OF SEQ ID NOS: 35
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 30
;; LENGTH: 121
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: Q48480-K. P
US-09-376-330-30

Query Match 7.3%; Score 121.5; DB 4; Length 121;
Best Local Similarity 28.3%; Pred. No. 1.8e-05;
Matches 36; Conservative 20; Mismatches 32; Indels 39; Gaps 5;

QY 96 DKVLYDPTDVLVRDGLKPLMDTDLGNNVVGACIDLFVEREGYKOKIGMA-----DG 147
DB 5 DKVFLDSTVYKADLGELLDVPLGNLVAAYKDIY---EGVKKFSASASDDGVPNG 61
QY 148 EX-----YFNACVLLINKKRRHIF-----KMSCWEVQYKDVMOYQOD 188
DB 62 EYLQKTLNNNPDEYFQAGIIVFNKQVVEENTFAELMRVLAKAKKY-----FLDQ 112
QY 189 DILNGLF 195
DB 113 DIMNKVF 119

RESULT 15
US-09-376-330-24
;; Sequence 24, Application US/09376330
;; Patent No. 6399321

;; GENERAL INFORMATION:
;; APPLICANT: Tessier, Daniel C.
;; APPLICANT: Dignard, Daniel
;; APPLICANT: Bergeron, John J.M.
;; APPLICANT: Thomas, David Y.
;; TITLE OF INVENTION: Method for screening for
;; TITLE OF INVENTION: UDP-glucose:glycoprotein glucosyltransferase (UGCT) activity
;; FILE REFERENCE: 2139-9"US"
;; CURRENT APPLICATION NUMBER: US/09/376,330
;; CURRENT FILING DATE: 1999-08-18
;; NUMBER OF SEQ ID NOS: 35
;; SOFTWARE: FastSeq for Windows Version 3.0
;; SEQ ID NO 24
;; LENGTH: 110
;; TYPE: PRT
;; ORGANISM: Artificial Sequence
;; FEATURE:
;; OTHER INFORMATION: GALT-S. typ
US-09-376-330-24

Query Match 6.9%; Score 115; DB 4; Length 110;
Best Local Similarity 28.8%; Pred. No. 7.7e-05;
Matches 32; Conservative 17; Mismatches 36; Indels 26; Gaps 4;

QY 97 KVLVLDPTDVLVRDGLKPLMDTDLGNNVVGACIDLFVEREGYKOKIG 143
DB 6 KVLVLDADVYCKGDISQLHLGLNGA-VAAVVKDVEPMQE--KAVSRLSDELLGQYFN 61
QY 144 MADGEYFNAGVLLINKKRRRHIFKMSCWEQYKDVMOYQODDILN 192
DB 57 LVSG--YFNAGFTLXIXPLMTAENISKAIEMLKQBEVVQRTTHDDQVFN 105

Search completed: December 2, 2002, 12:03:14
Job time : 12.3901 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 12:00:15 ; Search time 6.13314 Seconds
(without alignments)
794.504 Million cell updates/sec

Title: US-10-007-267-4

Perfect score: 1657

Sequence: 1 MDIVFAADNYKAVYLCVAAK.....KCMQWRKKLSARFLRKIV 306

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubppaa/US08_NEW_PUB.pep:*
2: /cgn2_6/ptodata/1/pubppaa/PCT_NEW_PUB.pep:*
3: /cgn2_6/ptodata/1/pubppaa/US06_NEW_PUB.pep:*
4: /cgn2_6/ptodata/1/pubppaa/US07_NEW_PUB.pep:*
5: /cgn2_6/ptodata/1/pubppaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubppaa/PCTUS_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubppaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubppaa/US09_NEW_PUB.pep:*
9: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
10: /cgn2_6/ptodata/1/pubppaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubppaa/US10_NEW_PUB.pep:*
12: /cgn2_6/ptodata/1/pubppaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubppaa/US60_NEW_PUB.pep:*
14: /cgn2_6/ptodata/1/pubppaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1657	100.0	306	12 US-10-007-267-4	Sequence 4, Appl1
2	176.5	10.7	674	10 US-09-765-272-200	Sequence 200, App
3	157	9.5	372	10 US-09-881-752A-320	Sequence 320, App
4	132	8.0	328	10 US-09-829-482-6	Sequence 6, Appl1
5	120.5	7.2	214	10 US-09-925-999-864	Sequence 864, App
6	118.5	7.2	371	9 US-09-992-598-171	Sequence 171, App
7	118.5	7.2	371	10 US-09-989-722-171	Sequence 171, App
8	118.5	7.2	371	10 US-09-989-723-171	Sequence 171, App
9	118.5	7.2	371	10 US-09-989-729-171	Sequence 171, App
10	118.5	7.2	371	10 US-09-989-727-171	Sequence 171, App
11	118.5	7.2	371	10 US-09-989-731-171	Sequence 171, App
12	118.5	7.2	371	10 US-09-989-732-171	Sequence 171, App
13	118.5	7.2	371	10 US-09-991-073-171	Sequence 171, App
14	118.5	7.2	371	10 US-09-990-442-171	Sequence 171, App
15	118.5	7.2	371	10 US-09-991-163-171	Sequence 171, App
16	118.5	7.2	371	10 US-09-993-604-171	Sequence 171, App
17	118.5	7.2	371	10 US-09-990-456-171	Sequence 171, App
18	118.5	7.2	371	10 US-09-989-721-171	Sequence 171, App
19	118.5	7.2	371	12 US-10-052-586-146	Sequence 146, App

20	118	7.1	281	9 US-09-895-913A-282	Sequence 282, App
21	98	5.9	346	10 US-09-829-482-2	Sequence 2, Appl1
22	87	5.3	394	12 US-10-036-507-2	Sequence 2, Appl1
23	82.5	5.0	756	10 US-09-745-763-142	Sequence 142, App
24	81	4.9	389	12 US-10-036-507-15	Sequence 15, Appl
25	81	4.9	389	12 US-10-007-693-112	Sequence 112, App
26	80.5	4.9	437	10 US-09-765-272-70	Sequence 70, Appl
27	78.5	4.7	593	10 US-09-815-242-12306	Sequence 12306, A
28	78.5	4.7	811	10 US-09-835-382-32	Sequence 32, Appl
29	78.5	4.7	1019	10 US-09-434-066-23	Sequence 23, Appl
30	78.5	4.7	2310	10 US-09-995-542-10	Sequence 10, Appl
31	78	4.7	566	10 US-09-829-482-5	Sequence 5, Appl1
32	77.5	4.7	536	10 US-09-815-242-12372	Sequence 12372, A
33	77.5	4.7	536	10 US-09-815-242-12978	Sequence 12978, A
34	77	4.6	311	9 US-09-981-876-191	Sequence 191, App
35	76.5	4.6	495	10 US-09-815-242-5040	Sequence 5040, Ap
36	76.5	4.6	1066	9 US-09-815-242-11716	Sequence 11716, A
37	76.5	4.6	1066	9 US-09-423-126-3	Sequence 3, Appl1
38	76.5	4.6	1066	10 US-09-280-197-5	Sequence 5, Appl1
39	74.5	4.5	316	10 US-09-765-272-156	Sequence 156, App
40	74	4.5	316	10 US-09-954-314-10	Sequence 10, Appl
41	74	4.5	416	9 US-09-971-536-60	Sequence 60, Appl
42	74	4.5	1169	10 US-09-771-161A-222	Sequence 222, App
43	73.5	4.4	333	10 US-09-829-482-3	Sequence 3, Appl1
44	73.5	4.4	342	9 US-09-911-150-4	Sequence 4, Appl1
45	73.5	4.4	493	10 US-09-815-242-12052	Sequence 12052, A

ALIGNMENTS

RESULT 1
US-10-007-267-4
Sequence 4, Application US/10007267
Patent No. US20020127682A1
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klausner & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 4:
SEQUENCE CHARACTERISTICS:
LENGTH: 306 amino acids
TYPE: amino acid

```

;          TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 4:
US-10-007-267-4

```

Query Match	100.0%;	Score 1657;	DB 12;	Length 306;
Best Local Similarity	100.0%;	Pred. No. 6.2e-161;		
Matches 306;	Conservative 0;	Mismatches 0;	Indels 0;	Gaps 0;

Qy	1	MDIYFAADNDNAAALCYAAKSV	EAHNDTEIRFVLDAGISEENRA	VAANLGGGINIR	60	
Db	1	MDIYFAADNDNAAALCYAAKSV	EAHNDTEIRFVLDAGISEENRA	VAANLGGGINIR	60	
Qy	61	IDVNPEDRPA	PLNIRHSTITTYARLKG	EYIADCDKVL	YLDLPDYLVRDGLKPLMTDGL	120
Db	61	IDVNPEDRPA	PLNIRHSTITTYARLKG	EYIADCDKVL	YLDLPDYLVRDGLKPLMTDGL	120
Qy	121	GNWVGACIDLFVEROE	GKORIGNADEYEFYFNAGVLL	INIKKMRHDI	IFMSCENEWOYK	180
Db	121	GNWVGACIDLFVEROE	GKORIGNADEYEFYFNAGVLL	INIKKMRHDI	IFMSCENEWOYK	180
Qy	181	DVMQYODDILNGLFK	GGVCTANRFRNMPNTNA	FMANGFASRHTD	PLYDTRNTAPVA	240
Db	181	DVMQYODDILNGLFK	GGVCTANRFRNMPNTNA	FMANGFASRHTD	PLYDTRNTAPVA	240
Qy	241	VSHYCGSAKPMHRDCT	WGAERFETELAGSLT	YVEEARGKLA	VPRTKMLORRKKLSAR	3000
Db	241	VSHYCGSAKPMHRDCT	WGAERFETELAGSLT	YVEEARGKLA	VPRTKMLORRKKLSAR	3000
Qy	301	FLRKTY	306			
Db	301	FLRKTY	306			

RESULT 2
US-09-765-272-200
; Sequence 200, Application US/09765272

```

;          TOPOLOGY: linear
;          MOLECULE TYPE: protein
;          SEQUENCE DESCRIPTION: SEQ ID NO: 200:
US-09-765-272-200

```

Query Match	10.7%;	Score 176.5;	DB 10;	Length 674;
Best Local Similarity	25.1%;	Pred. No. 7.2e-10;		
Matches	72;	Conservative 43;	Mismatches 107;	Indels 65; Gaps 11;

```
QY      3 IFFAADDDVVAAYALCYAANSAVSAAAPDTEIRHVVLDADAISEENRAAVALNAGGNIREFID 62
        ||||| : : : : : ||||| : : : : : ||||| : : : : : ||||| : : : : :
Db      289 IYLAANYGVVDVQLTTIKSI--CCHNRSIRFYLHSDPME-----WIKOLN 327

QY      63 VAPEDFAFPPLNIR-----HSITTYAALKGEIYIADCDKLYLIDTVLVPRDL 111
        | | | | | : : : : : | | | | | : : : : : | | | | | : : : : :
Db      328 KRLKEFDSEIIINCRATSEIOISCKSDISITYVELFYFLAFDAQE-DKALYIDLCDLVVTKNL 386

QY      112 KPLMTDTDEGMNVGACIDLFVERBEGYKOKIGMA-DEGYFNAGVLLILNKRRRHDIK 170
        || : || : || : || : || : || : || : || : || : || : || : || : || : || :
Db      387 DDLFTDTLODPRIAAVRPG-----GRAYFOEIFMAGVLLVNNNAFMKKEMTQ 435

QY      171 MSCENEQKYDMVOYODDILINGLFKGVCGYANSRENEMPTNTAEMANGASRHHDTPLYL 230
        : : : : : : : : : || || || || : : : : : : : : : : : : : : : : : :
Db      436 KLIDVTNEWHEDVDQAQSILIMTLFEHKWLELTDVDYHNIVYHKOFADYQURPEGODYR-- 492

QY      231 DKTNTPMPAVASHYGSGAKPV-----HRCSJTW----GAERTEL 266
        | : | : | : | : | : | : | : | : | : | : | : | : | : | : | : | :
Db      493 -----AIHHYLSHRPKMDLAOQYRE-VWWYYHGIE-WTEL 527
```

RESULT 3
US-09-881-752A-320
; Sequence 320, Application US/09881752A
; Patent No. US20020115078A1

Query Match	9.5%	Score 157;	DB 10;	Length 372;
Best Local Similarity	20.4%;	Pred. No. 3e-08;		
Matches 75; Conservative	53;	Mismatches 127;	Indels 112;	Gaps 15

```

QY 3 IYEADDNDNVAALCYAAKSVSEA---IPDDE-----IFPHYLDAGISEENAAVAAN 51
Db 7 IYIADPNHAMPAGVSLYSMLCAKTEHPQSDNSEKILFYIHCILVNDLSLENDOSKIKET 66
QY 52 L---RCGGINREIDVNPEDFAGFP-----LNIHISITTYA-----RLTKIGE 90
Db 67 LAPFSAFSSLEFLDISTRPLNATPIEBSAIDKINEAFIQLIYIAKTSKMWKMRCLFLAS 126
QY 91 YIADCDKVLVLETDVLVBDGLPLMDDTDLGNGWVAGACIDLEVEEROEQKOKIGMAD--- 147
Db 127 LEPQYDIIIMFPADLFLINDVSESEFFIPLDQYGYGAAMDFASDSPNHQFQVREKDDPQA 166
QY 148 ---EYY-----FNAGVLLINLKKWRHDIKFKSCMEVBEQYKDVMOYODDI 190

```

DB 187 FSLYEHYNESDMQIYESNNYAGFLVYNLKLMDRADHLEERLLNLTHQGCQCFYEPDOL 246
QY 191 INGLFKGVCYANSRFPNMP--TNVAFMANGFASRHTDPLTLDPTNAMPVAVSHYGS 247
DB 247 LT-----LACY--QKVLIPFYIYNTHPRMAN--OKRFP--DKKE-----IVMLHFTV 289
QY 248 AKPMWRDCTVWGAEFTELASLTTPVEWRGLAVPPTKMLQWRKKL-----S 298
DB 290 GKPW-----VLPTFSYSKEMHETLTKTPRYAEXS 318
QY 299 ARFLRKI 305
DB 319 VKFLKQM 325

RESULT 4
US-09-829-482-6
Sequence 6, Application US/09829482
Patent No. US2002001843A1
GENERAL INFORMATION:
APPLICANT: LIGHTNER, JONATHAN EDWARD
TITLE OF INVENTION: STARCH BIOSYNTHETIC ENZYMES
NUMBER OF SEQUENCES: 26
CORRESPONDENCE ADDRESS:
ADDRESSEE: E. I. DU PONT DE NEMOURS AND COMPANY
STREET: 1007 MARKET STREET
CITY: WILMINGTON
STATE: DELAWARE
COUNTRY: U.S.A.
ZIP: 19898
COMPUTER READABLE FORM:
MEDIUM TYPE: DISKETTE, 3.50 INCH
COMPUTER: IBM PC COMPATIBLE
OPERATING SYSTEM: MICROSOFT WINDOWS 95
SOFTWARE: MICROSOFT WORD VERSION 7.0A
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/829,482
FILING DATE: 10-APR-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/852615
FILING DATE: MAY 7, 1997
ATTORNEY/AGENT INFORMATION:
NAME: MAJARIAN, WILLIAM R.
REGISTRATION NUMBER: 41,173
REFERENCE/DOCKET NUMBER: BB-1083-B
TELEPHONE: 302-992-4926
TELEFAX: 302-773-0164
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 328 amino acids
TYPE: amino acid
STRANDEDNESS: not relevant
TOPOLOGY: linear
MOLECULE TYPE: peptide
SEQUENCE DESCRIPTION: SEQ ID NO: 6:
US-09-829-482-6

Query Match 8.0%; Score 132; DB 10; Length 328;
Best Local Similarity 23.1%; Pred. No. 8.9e-06;
Matches 56; Conservative 42; Mismatches 64; Indels 80; Gaps 15;

QY 80 ITTYARKIGETIADCKVLYIDTVLRDGLKPLMDTDLGNNWYGACIDLFVER----- 134
DB 95 VINYSLRLTMEFV-EXERVYLDADIQVFDNDHFLDLDKGFY--AVKDCFCERTWST 151
QY 135 ---QGYKO---KITMADGE-----YYNAGVLLINLKKRRHDFRMSCEWQYKD 181
DB 152 PPDYDGYCQQRDEYAVMRELRGPPPLTFNAGMFV-----HE-----PGIGTAKD 197
QY 182 VMO-----YODDILNGLFKGVCYANSRFPNMPNTNVAFMANGFASRHTDPLYLDR 232

DB 198 LLDALVTPPTFFAODFLNMFRR-----EQYKPIPVNVLV-AMLMRHPENVDDQ 249
QY 223 TTNAPRVAVSHYGS-AKPMWRDCTVWGAEFTELASLTTPVEWRGLAVPPTKMLQ 291
DB 250 -----VKVHYCAAGSKPW-----RFT-----GKEENNRE-----DIKMLVK 282
QY 292 RW 293
DB 283 RW 284

RESULT 5
US-09-925-299-864
Sequence 864, Application US/09925299
Patent No. US20020055627A1
GENERAL INFORMATION:
APPLICANT: Rosen et al.
TITLE OF INVENTION: Nucleic Acids, Proteins and Antibodies
FILE REFERENCE: PA102
CURRENT APPLICATION NUMBER: US/09/925,299
PRIOR FILING DATE: 2001-08-10
PRIOR APPLICATION NUMBER: PCT/US00/05883
PRIOR FILING DATE: 2000-03-08
PRIOR APPLICATION NUMBER: 60/124,270
PRIOR FILING DATE: 1999-03-12
NUMBER OF SEQ ID NOS: 1556
SOFTWARE: PatentIn Ver. 2.0
SEQ ID NO 864
LENGTH: 214
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
NAME/KEY: SITE
LOCATION: (138)
OTHER INFORMATION: Xaa equals any of the naturally occurring L-amino acids
US-09-925-299-864

Query Match 7.3%; Score 120.5; DB 10; Length 214;
Best Local Similarity 24.5%; Pred. No. 7.3e-05;
Matches 47; Conservative 35; Mismatches 69; Indels 41; Gaps 10;

QY 97 KVLVLDTVLRDGLKPLMDTDLGNNWYGACID-----LFVEROEG-----YK 139
DB 16 KVIYDDDVYQGDIGELDTDTLALGHAASFDDDDLSAQDINLTVGLQNTYMYGLDYR 75
QY 140 OK-----IGMADGEYFNAGVLLINLKKRRHDFRMSCEWQYKDVMQYODDILNGLF 195
DB 76 KKAIRDGLISPTCSFNGVIYANNTEWKHORITKOLEKMQ--KNV-----EENLYSSL 129
QY 196 KGCVCYANSRFPNMPNTNVAFMANGFASRH---TPPLVDRNTNAMPVAVSHYGSARKPW 251
DB 130 GGGVATSPYLIVF-HGKYSTINPLWHLIRLGNPDRYSE--HFLQEAKLHMGNRHHPW 186
QY 252 -----HRDCTW 258
DB 187 DEPSVHND--LW 196

RESULT 6
US-09-992-598-171
Sequence 171, Application US/09992598
Patent No. US20020160384A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Bolstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerritsen, Mary E.

APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730PLC20
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090445
PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090472
PRIOR FILING DATE: 1998-06-24

PRIOR APPLICATION NUMBER: 60/088212	PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217	PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655	PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734	PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088738	PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742	PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810	PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824	PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088828	PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088855	PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088861	PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876	PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105	PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089106	PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440	PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512	PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514	PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089533	PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538	PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598	PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599	PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089610	PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089611	PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089655	PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089801	PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948	PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948	PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952	PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090254	PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252	PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349	PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355	PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090431	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090435	PRIOR FILING DATE: 1998-06-24
PRIOR APPLICATION NUMBER: 60/090444	PRIOR FILING DATE: 1998-06-24

1	PRIOR FILING DATE: 1998-06-24	
2	PRIOR APPLICATION NUMBER: 60/090445	
3	PRIOR FILING DATE: 1998-06-24	
4	PRIOR APPLICATION NUMBER: 60/090472	
5	PRIOR FILING DATE: 1998-06-24	
6	PRIOR APPLICATION NUMBER: 60/090535	
7	PRIOR FILING DATE: 1998-06-24	
8	PRIOR APPLICATION NUMBER: 60/090540	
9	PRIOR FILING DATE: 1998-06-24	
10	PRIOR APPLICATION NUMBER: 60/090542	
11	PRIOR FILING DATE: 1998-06-24	
12	PRIOR APPLICATION NUMBER: 60/090557	
13	PRIOR FILING DATE: 1998-06-24	
14	PRIOR APPLICATION NUMBER: 60/090676	
15	PRIOR FILING DATE: 1998-06-25	
16	PRIOR APPLICATION NUMBER: 60/090678	
17	PRIOR FILING DATE: 1998-06-25	
18	PRIOR APPLICATION NUMBER: 60/090690	
19	PRIOR FILING DATE: 1998-06-25	
20	PRIOR APPLICATION NUMBER: 60/090694	
21	PRIOR FILING DATE: 1998-06-25	
22	PRIOR APPLICATION NUMBER: 60/090695	
23	PRIOR FILING DATE: 1998-06-25	
24	PRIOR APPLICATION NUMBER: 60/090696	
25	PRIOR FILING DATE: 1998-06-25	
26	PRIOR APPLICATION NUMBER: 60/090862	
27	PRIOR FILING DATE: 1998-06-26	
28	PRIOR APPLICATION NUMBER: 60/090863	
29	PRIOR FILING DATE: 1998-06-26	
30	PRIOR APPLICATION NUMBER: 60/091360	
31	PRIOR FILING DATE: 1998-07-01	
32	PRIOR APPLICATION NUMBER: 60/091478	
33	PRIOR FILING DATE: 1998-07-02	
34	PRIOR APPLICATION NUMBER: 60/091544	
35	PRIOR FILING DATE: 1998-07-01	
36	PRIOR APPLICATION NUMBER: 60/091519	
37	PRIOR FILING DATE: 1998-07-02	
38	PRIOR APPLICATION NUMBER: 60/091626	
39	PRIOR FILING DATE: 1998-07-02	
40	PRIOR APPLICATION NUMBER: 60/091633	
41	PRIOR FILING DATE: 1998-07-02	
42	PRIOR APPLICATION NUMBER: 60/091978	
43	PRIOR FILING DATE: 1998-07-07	
44	PRIOR APPLICATION NUMBER: 60/091982	
45	PRIOR FILING DATE: 1998-07-07	
46	PRIOR APPLICATION NUMBER: 60/092182	
47	PRIOR FILING DATE: 1998-07-09	

Query Match	7.2%;	Score 118.5;	DB 10;	Length 371;
Best Local Similarity	20.4%;	Pred. No. 0.00025;		
Matches 48;	Conservative 38;	Mismatches 74;	Indels 75;	Gaps 9;

Qy	82	TYAKLKGEYIADCDKVLIEDTVLVDGLKPLMDTL-----	119
Db	151	TFAFYLPILPSSAKKAIYVGGDILATYNTLKRGHAAFSEDCDASAKVIR	210
Qy	120	--GG--MVGACIDLFVPROBGYKOKIGMADGEYYPAGVLLINLKRRRHDIFKSCSEW	175
Db	211	GAGQNYNIG-----YLDKKERYRKLSTMASTCSFPPGVFVANLPEWKQNTITNOLEKW	265
Qy	176	VEQRYDWQYODDILNGLFKGVCYVANSFNEPMTNYAFMANGFASRHT--DPLYLR-	232
Db	266	MK-----LNVESGL-----YSRLTAGSITPPRLIIV--FYQGHSTIDPMNVHR	307
Qy	233	-----TNTMPPAVASHICGSAAKPFMRHDCVWGAEKFTELAGSLITVPEEW	277
Db	308	LGSAGAKRYSEQFVKAAKLLHMNHKLPMWR-----TASYDPEWKKW	349

RESULT 8
US-09-989-723-171
; Sequence 171, Application US/09989723
; Patent No. US20020072092A1

GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Bolstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Geritsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C62
CURRENT APPLICATION NUMBER: US/09/989,723
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326

PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090252
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090254
PRIOR FILING DATE: 1998-06-22
PRIOR APPLICATION NUMBER: 60/090349
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090355
PRIOR FILING DATE: 1998-06-23
PRIOR APPLICATION NUMBER: 60/090429
PRIOR FILING DATE: 1998-06-24

```

; PRIOR APPLICATION NUMBER: 60/090431
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090435
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090444
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090445
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090472
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090535
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090540
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090542
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090557
; PRIOR FILING DATE: 1998-06-24
; PRIOR APPLICATION NUMBER: 60/090676
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090678
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090690
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090694
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090695
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090696
; PRIOR FILING DATE: 1998-06-25
; PRIOR APPLICATION NUMBER: 60/090862
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/090863
; PRIOR FILING DATE: 1998-06-26
; PRIOR APPLICATION NUMBER: 60/091360
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091478
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

Query Match          7.2%  Score 118.5;  DB 10;  Length 371;
Best Local Similarity 20.4%;  Pred. No. 0.00025;
Matches 48;  Conservative 38;  Mismatches 74;  Indels 75;  Gaps 9;

QY 82 TYARLKGIVYIDCKVYLDVYRDLKPLMDTL----- 119
DB 151 TTRREYPLIVLSAKKATYMDDDVYOGDILALYNALKPGHAAAFSEDCDSASTKRVIR 210
QY 120 --GG--NMWACIDLFVERQESYKOKIGMADEYFYNAGVLLINLKMRHRDIFKMSCEM 175
DB 211 GAGNGQNYNG-----YLDKKRIRKLSKASTCSNPGVFYANLTFMKRQMITNOLEKW 265
QY 176 VEOYKDVMOYODDILNGLFGKGVCVANSRNFNPMPTNAFMKNGFASRHT--DPLYLDR- 232
DB 266 MK-----LNVEEGD-----YSRLAGSITTPPLLIV--FYQSHSTIDPMNVNRH 307
QY 233 -----TNTAMPVAVSHYCGSAKPKPHRDCIYWGAEFTELGLASTTVEEEM 277
DB 308 LGSAGKRYSPQFVAKAKLLHWNHGLKPMGR-----TASYTDVWEKW 349
```

```

RESULT 9
US-09-989-279-171
; Sequence 171, Application US/09989279
; Patent No. US20020072496A1
; GENERAL INFORMATION:
; APPLICATION: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paonli, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730P1C56
; CURRENT APPLICATION NUMBER: US/09/989, 279
; PRIOR FILING DATE: 2001-11-19
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910
; PRIOR FILING DATE: 1998-03-20
; PRIOR APPLICATION NUMBER: 60/083322
; PRIOR FILING DATE: 1998-04-28
; PRIOR APPLICATION NUMBER: 60/084600
; PRIOR FILING DATE: 1998-05-07
; PRIOR APPLICATION NUMBER: 60/087106
; PRIOR FILING DATE: 1998-05-28
; PRIOR APPLICATION NUMBER: 60/087607
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087609
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087759
; PRIOR FILING DATE: 1998-06-02
; PRIOR APPLICATION NUMBER: 60/087827
; PRIOR FILING DATE: 1998-06-03
; PRIOR APPLICATION NUMBER: 60/088021
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088025
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088026
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088028
; PRIOR FILING DATE: 1998-06-04
; PRIOR APPLICATION NUMBER: 60/088029
; PRIOR FILING DATE: 1998-06-04
```


OY 233 -----TNTAMPVAVSHYCSAKPMHROCTVWGAEFFTELASLTTPPEEM 277
DB 308 LGSSACKRYSPOFVKAKLHMNGHLKPMGR-----TASYDVMWKW 349
RESULT 10
US-09-989-727-171
Sequence 171, Application US/09989727
Patent No. US20020072497A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Guirney, Austin L.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C65
CURRENT APPLICATION NUMBER: US/09/989,727
PRIOR FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026

PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089908
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089947
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089948
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/089952
PRIOR FILING DATE: 1998-06-19
PRIOR APPLICATION NUMBER: 60/090246
PRIOR FILING DATE: 1998-06-22

PROR APPLICATION NUMBER: 60/088021
PROR FILING DATE: 1998-06-04
PROR APPLICATION NUMBER: 60/088025
PROR FILING DATE: 1998-06-04
PROR APPLICATION NUMBER: 60/088026
PROR FILING DATE: 1998-06-04
PROR APPLICATION NUMBER: 60/088028
PROR FILING DATE: 1998-06-04
PROR APPLICATION NUMBER: 60/088029
PROR FILING DATE: 1998-06-04
PROR APPLICATION NUMBER: 60/088030
PROR FILING DATE: 1998-06-04
PROR APPLICATION NUMBER: 60/088033
PROR FILING DATE: 1998-06-04
PROR APPLICATION NUMBER: 60/088326
PROR FILING DATE: 1998-06-04
PROR APPLICATION NUMBER: 60/088167
PROR FILING DATE: 1998-06-05
PROR APPLICATION NUMBER: 60/088202
PROR FILING DATE: 1998-06-05
PROR APPLICATION NUMBER: 60/088212
PROR FILING DATE: 1998-06-05
PROR APPLICATION NUMBER: 60/088217
PROR FILING DATE: 1998-06-05
PROR APPLICATION NUMBER: 60/088655
PROR FILING DATE: 1998-06-09
PROR APPLICATION NUMBER: 60/088734
PROR FILING DATE: 1998-06-10
PROR APPLICATION NUMBER: 60/088738
PROR FILING DATE: 1998-06-10
PROR APPLICATION NUMBER: 60/088742
PROR FILING DATE: 1998-06-10
PROR APPLICATION NUMBER: 60/088810
PROR FILING DATE: 1998-06-10
PROR APPLICATION NUMBER: 60/088824
PROR FILING DATE: 1998-06-10
PROR APPLICATION NUMBER: 60/088826
PROR FILING DATE: 1998-06-10
PROR APPLICATION NUMBER: 60/088858
PROR FILING DATE: 1998-06-11
PROR APPLICATION NUMBER: 60/088861
PROR FILING DATE: 1998-06-11
PROR APPLICATION NUMBER: 60/088876
PROR FILING DATE: 1998-06-11
PROR APPLICATION NUMBER: 60/089105
PROR FILING DATE: 1998-06-12
PROR APPLICATION NUMBER: 60/089440
PROR FILING DATE: 1998-06-16
PROR APPLICATION NUMBER: 60/089512
PROR FILING DATE: 1998-06-16
PROR APPLICATION NUMBER: 60/089514
PROR FILING DATE: 1998-06-16
PROR APPLICATION NUMBER: 60/089532
PROR FILING DATE: 1998-06-17
PROR APPLICATION NUMBER: 60/089538
PROR FILING DATE: 1998-06-17
PROR APPLICATION NUMBER: 60/089598
PROR FILING DATE: 1998-06-17
PROR APPLICATION NUMBER: 60/089599
PROR FILING DATE: 1998-06-17
PROR APPLICATION NUMBER: 60/089600
PROR FILING DATE: 1998-06-17
PROR APPLICATION NUMBER: 60/089653
PROR FILING DATE: 1998-06-17
PROR APPLICATION NUMBER: 60/089801
PROR FILING DATE: 1998-06-18
PROR APPLICATION NUMBER: 60/089907
PROR FILING DATE: 1998-06-18
PROR APPLICATION NUMBER: 60/089908
PROR FILING DATE: 1998-06-18
PROR APPLICATION NUMBER: 60/089947
PROR FILING DATE: 1998-06-19
PROR APPLICATION NUMBER: 60/089948

PROR FILING DATE: 1998-06-19
PROR APPLICATION NUMBER: 60/089952
PROR FILING DATE: 1998-06-19
PROR APPLICATION NUMBER: 60/090246
PROR FILING DATE: 1998-06-22
PROR APPLICATION NUMBER: 60/090252
PROR FILING DATE: 1998-06-22
PROR APPLICATION NUMBER: 60/090254
PROR FILING DATE: 1998-06-22
PROR APPLICATION NUMBER: 60/090349
PROR FILING DATE: 1998-06-23
PROR APPLICATION NUMBER: 60/090355
PROR FILING DATE: 1998-06-23
PROR APPLICATION NUMBER: 60/090429
PROR FILING DATE: 1998-06-24
PROR APPLICATION NUMBER: 60/090431
PROR FILING DATE: 1998-06-24
PROR APPLICATION NUMBER: 60/090435
PROR FILING DATE: 1998-06-24
PROR APPLICATION NUMBER: 60/090444
PROR FILING DATE: 1998-06-24
PROR APPLICATION NUMBER: 60/090445
PROR FILING DATE: 1998-06-24
PROR APPLICATION NUMBER: 60/090472
PROR FILING DATE: 1998-06-24
PROR APPLICATION NUMBER: 60/090535
PROR FILING DATE: 1998-06-24
PROR APPLICATION NUMBER: 60/090540
PROR FILING DATE: 1998-06-24
PROR APPLICATION NUMBER: 60/090542
PROR FILING DATE: 1998-06-24
PROR APPLICATION NUMBER: 60/090557
PROR FILING DATE: 1998-06-24
PROR APPLICATION NUMBER: 60/090676
PROR FILING DATE: 1998-06-25
PROR APPLICATION NUMBER: 60/090678
PROR FILING DATE: 1998-06-25
PROR APPLICATION NUMBER: 60/090690
PROR FILING DATE: 1998-06-25
PROR APPLICATION NUMBER: 60/090694
PROR FILING DATE: 1998-06-25
PROR APPLICATION NUMBER: 60/090695
PROR FILING DATE: 1998-06-25
PROR APPLICATION NUMBER: 60/090696
PROR FILING DATE: 1998-06-25
PROR APPLICATION NUMBER: 60/090862
PROR FILING DATE: 1998-06-26
PROR APPLICATION NUMBER: 60/090863
PROR FILING DATE: 1998-06-26
PROR APPLICATION NUMBER: 60/091360
PROR FILING DATE: 1998-07-01
PROR APPLICATION NUMBER: 60/091478
PROR FILING DATE: 1998-07-02
PROR APPLICATION NUMBER: 60/091544
PROR FILING DATE: 1998-07-01
PROR APPLICATION NUMBER: 60/091519
PROR FILING DATE: 1998-07-02
PROR APPLICATION NUMBER: 60/091626
PROR FILING DATE: 1998-07-02
PROR APPLICATION NUMBER: 60/091633
PROR FILING DATE: 1998-07-02
PROR APPLICATION NUMBER: 60/091978
PROR FILING DATE: 1998-07-07
PROR APPLICATION NUMBER: 60/091982
PROR FILING DATE: 1998-07-07
PROR APPLICATION NUMBER: 60/092182
PROR FILING DATE: 1998-07-09

Query Match 7.2% Score 118.5; DB 10; Length 371;

Best Local Similarity 20.4% Pred. No. 0.00025;
Matches 48; Conservative 38; Mismatches 74; Indels 75; Gaps 9;

QY 82 TYARLGEYIADCDKVLXLDITDVIVRDGLKPLMDTDL----- 119

Db 151 TFAEFYLLVPSAKAIYMDDDVIVCGDILALYNTALKPGHAAAFSEDCDSASTKVIR 210
Oy 120 --GG--MNVGACIDLFVERQEGYKOKIGMADCEYFENAGVLLINKKRHRHIFKSCSEW 175
Db 211 GAGNGYNYIG----YLDYKKERIKLSKASTCSFNGCVFANLEKRONITMOLKRW 265
Oy 176 VEQYVDVQYODODILNGLFKGVCYANSRNFNPTNYAFMANGFASRHT--DPLYLDR- 232
Db 266 MK-----LWVEBGL-----XSRTLAGSITTPPLIIV--FYQCHSTIDPMMNRRH 307
Oy 233 -----TNTAMPVAVSHYCGSAKPMWRDCTVWGAERFTLAGSLTTVPEEW 277
Db 308 LGSAGKRRSPQFVKAAKLLHWNHGLKPMGR-----TASYTDVWEKW 349
RESULT 12
US-09-989-732-171
Sequence 171, Application US/09989732
Patent No. US20020123463A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Bolstein, David
APPLICANT: Desnoyers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gerlitsen, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kljavin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zemin
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
TITLE OF INVENTION: Acids Encoding the Same
FILE REFERENCE: P2730P1C37
CURRENT APPLICATION NUMBER: US/09/989,732
CURRENT FILING DATE: 2001-11-19
PRIOR APPLICATION NUMBER: 60/04987
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/075945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07
PRIOR APPLICATION NUMBER: 60/087106
PRIOR FILING DATE: 1998-05-28
PRIOR APPLICATION NUMBER: 60/087607
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087609

PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087759
PRIOR FILING DATE: 1998-06-02
PRIOR APPLICATION NUMBER: 60/087827
PRIOR FILING DATE: 1998-06-03
PRIOR APPLICATION NUMBER: 60/088021
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088025
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088026
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088028
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088029
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088030
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088033
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088326
PRIOR FILING DATE: 1998-06-04
PRIOR APPLICATION NUMBER: 60/088167
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088202
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088212
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088217
PRIOR FILING DATE: 1998-06-05
PRIOR APPLICATION NUMBER: 60/088655
PRIOR FILING DATE: 1998-06-09
PRIOR APPLICATION NUMBER: 60/088734
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088738
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088742
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088810
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088824
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088826
PRIOR FILING DATE: 1998-06-10
PRIOR APPLICATION NUMBER: 60/088858
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088861
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/088876
PRIOR FILING DATE: 1998-06-11
PRIOR APPLICATION NUMBER: 60/089105
PRIOR FILING DATE: 1998-06-12
PRIOR APPLICATION NUMBER: 60/089440
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089512
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089514
PRIOR FILING DATE: 1998-06-16
PRIOR APPLICATION NUMBER: 60/089532
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089538
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089598
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089599
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089600
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089653
PRIOR FILING DATE: 1998-06-17
PRIOR APPLICATION NUMBER: 60/089801
PRIOR FILING DATE: 1998-06-18
PRIOR APPLICATION NUMBER: 60/089907
PRIOR FILING DATE: 1998-06-18

```

Query Match      7.2%: Score 118.5; DB 10; Length 371;
Best Local Similarity 20.4%: Pred. No. 0.00025;
Matches 48; Conservative 38; Mismatches 74; Indels 75; Gaps 9;

QY      82  TYARKLGEYIADDDKVLVLDTPDVLVVRGLKPLMDTDL----- 119
      |::| |::| |::| |::| |::| |
Db      151  TFAHFYFLPLVPSAKAIYMDDDVIVGGDILALYNTALKPGHAAAFSEDCDSASTKVIAR 210
      |::| |::| |::| |::| |::| |

QY      120  --GG--NMWACIDLEFVERQEGYKOKITGMADGEYIFNAGVLLINLKKRRHIDIFKMSCW 175
      |::| |::| |::| |::| |::| |
Db      211  GAGQNYNYIG-----YLDKKERKIRKLMSKASTCSFNGVFVANTLEWKRONITQLEKW 265
      |::| |::| |::| |::| |::| |

QY      176  VEQYKDWQVQDDDLNLNLFKGGVCYANSRFNMPNTNYAFMANGFASRHT--DPLYLDR- 232
      ::| |::| |::| |::| |::| |
Db      266  MK-----LNVEGL-----YSTLGLSITTPLLIV--FYQHSITDPMNVRH 307
      ::| |::| |::| |::| |::| |

QY      233  -----TNTAPVAVSHYCSGSAKPMHRDCTVWGAERFTELGLSTVPEEW 277
      :| | | | | | |
Db      308  LGSAGKRYSPQFVKAALHMHNGHLKRWGR-----TASTVDWEKW 349
      :| | | | | | |

RESULT 13
US-09-991-073-171
Sequence 171, Application US/09991073
Patent No. US20020127576A1
GENERAL INFORMATION:
APPLICANT: Ashkenazi, Avi J.
APPLICANT: Baker, Kevin P.
APPLICANT: Botstein, David
APPLICANT: Desnovers, Luc
APPLICANT: Eaton, Dan L.
APPLICANT: Ferrara, Napoleone
APPLICANT: Fong, Sherman
APPLICANT: Gerber, Hanspeter
APPLICANT: Gertsens, Mary E.
APPLICANT: Goddard, Audrey
APPLICANT: Godowski, Paul J.
APPLICANT: Grimaldi, J. Christopher
APPLICANT: Gurney, Austin L.
APPLICANT: Kijavlin, Ivar J.
APPLICANT: Napier, Mary A.
APPLICANT: Pan, James
APPLICANT: Paoni, Nicholas F.
APPLICANT: Roy, Margaret Ann
APPLICANT: Stewart, Timothy A.
APPLICANT: Tumas, Daniel
APPLICANT: Watanabe, Colin K.
APPLICANT: Williams, P. Mickey
APPLICANT: Wood, William I.
APPLICANT: Zhang, Zhenli
TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
FILE REFERENCE: P2730P1C15
CURRENT APPLICATION NUMBER: US/09/991,073
CURRENT FILING DATE: 2001-11-14
PRIOR APPLICATION NUMBER: 60/049787
PRIOR FILING DATE: 1997-06-16
PRIOR APPLICATION NUMBER: 60/062250
PRIOR FILING DATE: 1997-10-17
PRIOR APPLICATION NUMBER: 60/065186
PRIOR FILING DATE: 1997-11-12
PRIOR APPLICATION NUMBER: 60/065311
PRIOR FILING DATE: 1997-11-13
PRIOR APPLICATION NUMBER: 60/066770
PRIOR FILING DATE: 1997-11-24
PRIOR APPLICATION NUMBER: 60/079945
PRIOR FILING DATE: 1998-02-25
PRIOR APPLICATION NUMBER: 60/078910
PRIOR FILING DATE: 1998-03-20
PRIOR APPLICATION NUMBER: 60/083322
PRIOR FILING DATE: 1998-04-28
PRIOR APPLICATION NUMBER: 60/084600
PRIOR FILING DATE: 1998-05-07

```

[illegible]

;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/092182
;; PRIOR FILING DATE: 1998-07-09

Query Match 7.2%; Score 118.5; DB 10; Length 371;
Best Local Similarity 20.4%; Pred. No. 0.00025;
Matches 48; Conservative 38; Mismatches 74; Indels 75; Gaps 9;

QY 82 TYARLKGIVYADCKVLYLDTVLVROSLKPLMDRL----- 119
DB 151 TFAHYLPPLVPSAKALYMDVYQGDILALYNTALPGHAAFESECDASIKVIR 210
QY 120 --GG--NWVACIDLFVEHOGSEYKQKIGMADEYFFNAGVLLINKKRRHDFKMSCFM 175
DB 211 GAGNGYNYTG-----YLDYKKERIRLSMKASTCSFNPQVFVANLTKMKRONITNLEKW 265
QY 176 VEOYVDWMOYQDODILNGLFKGVCYANSRFPMPNTNFMANGASRHT--DPLYLDR- 232
DB 266 MK-----LNVEEG-----YSRLAGSITTPPLLIY--FYQGHSTIDPMNVNRH 307
QY 233 -----TNTAMPVAASHYGSAGKPMWRDCTVWGAERFELAGSLTTPPEEW 277
DB 308 LGSSAGKRYSPQFVKAAKLHMNGHLKPMGR-----TASYTDVWEKW 349

RESULT 14
US-09-990-442-171
; Sequence 171, Application US/09990442
; Patent No. US20020132252A1
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary F.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kijavini, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC8
; CURRENT APPLICATION NUMBER: US/09/990,442
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13
; PRIOR APPLICATION NUMBER: 60/066770
; PRIOR FILING DATE: 1997-11-24
; PRIOR APPLICATION NUMBER: 60/075945
; PRIOR FILING DATE: 1998-02-25
; PRIOR APPLICATION NUMBER: 60/078910

;; PRIOR FILING DATE: 1998-03-20
;; PRIOR APPLICATION NUMBER: 60/083322
;; PRIOR FILING DATE: 1998-04-28
;; PRIOR APPLICATION NUMBER: 60/084600
;; PRIOR FILING DATE: 1998-05-07
;; PRIOR APPLICATION NUMBER: 60/087106
;; PRIOR FILING DATE: 1998-05-28
;; PRIOR APPLICATION NUMBER: 60/087607
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087609
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087759
;; PRIOR FILING DATE: 1998-06-02
;; PRIOR APPLICATION NUMBER: 60/087827
;; PRIOR FILING DATE: 1998-06-03
;; PRIOR APPLICATION NUMBER: 60/088021
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088025
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088026
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088028
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088029
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088030
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088033
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088326
;; PRIOR FILING DATE: 1998-06-04
;; PRIOR APPLICATION NUMBER: 60/088167
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088202
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088212
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088217
;; PRIOR FILING DATE: 1998-06-05
;; PRIOR APPLICATION NUMBER: 60/088655
;; PRIOR FILING DATE: 1998-06-09
;; PRIOR APPLICATION NUMBER: 60/088734
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088738
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088742
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088810
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088824
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088826
;; PRIOR FILING DATE: 1998-06-10
;; PRIOR APPLICATION NUMBER: 60/088858
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088861
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/088876
;; PRIOR FILING DATE: 1998-06-11
;; PRIOR APPLICATION NUMBER: 60/089105
;; PRIOR FILING DATE: 1998-06-12
;; PRIOR APPLICATION NUMBER: 60/089440
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089512
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089514
;; PRIOR FILING DATE: 1998-06-16
;; PRIOR APPLICATION NUMBER: 60/089532
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089538
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089598
;; PRIOR FILING DATE: 1998-06-17

;; PRIOR APPLICATION NUMBER: 60/089599
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089600
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089653
;; PRIOR FILING DATE: 1998-06-17
;; PRIOR APPLICATION NUMBER: 60/089801
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089907
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089908
;; PRIOR FILING DATE: 1998-06-18
;; PRIOR APPLICATION NUMBER: 60/089947
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/089948
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/089952
;; PRIOR FILING DATE: 1998-06-19
;; PRIOR APPLICATION NUMBER: 60/090246
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090252
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090254
;; PRIOR FILING DATE: 1998-06-22
;; PRIOR APPLICATION NUMBER: 60/090349
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090355
;; PRIOR FILING DATE: 1998-06-23
;; PRIOR APPLICATION NUMBER: 60/090429
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090431
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090435
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090444
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090445
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090472
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090535
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090540
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090542
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090557
;; PRIOR FILING DATE: 1998-06-24
;; PRIOR APPLICATION NUMBER: 60/090676
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090678
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090690
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090694
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090695
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090696
;; PRIOR FILING DATE: 1998-06-25
;; PRIOR APPLICATION NUMBER: 60/090862
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/090863
;; PRIOR FILING DATE: 1998-06-26
;; PRIOR APPLICATION NUMBER: 60/091360
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091478
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091544
;; PRIOR FILING DATE: 1998-07-01
;; PRIOR APPLICATION NUMBER: 60/091519
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091626

;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091633
;; PRIOR FILING DATE: 1998-07-02
;; PRIOR APPLICATION NUMBER: 60/091978
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/091982
;; PRIOR FILING DATE: 1998-07-07
;; PRIOR APPLICATION NUMBER: 60/092182
;; PRIOR FILING DATE: 1998-07-09

Query Match 7.2%; Score 118.5; DB 10; Length 371;
Best Local Similarity 20.4%; Pred. No. 0.00025;
Matches 48; Conservative 38; Mismatches 74; Indels 75; Gaps 9;

QY 82 TYARLKLGEYIADCKVLTLDIVLRDGLKPLMDL----- 119
DB 151 TFAREYPLIVPSAKKATVDDVIVQGDILALYNALPGHAAFESEDCDSATKVVIR 210
QY 120 --GG--NWGACIDLFVERQEGYKQKIGMADEYEFNAGVLLINLKKRRHDIKMSCEM 175
DB 211 GAGNOYNYIG-----YLDYKKERIRKLSKASTCSFNPGVFVANLTKWRONTNLEKW 265
QY 176 VEQYDVMQYODDILINGLFKGCVCYANSRENFMTNVAFMANGFASRHT--DPLYLDR- 232
DB 266 MK-----LNVEEGT-----YSRTLAGSIPTPLLIY--FYQCHSTIDPMNVHR 307
QY 233 -----TNTAMPVANSHYCGSAKPPHNRDCTVGAERFELAGSLTTPVEEM 277
DB 308 LGSSAGKRRYSPQFVRAAKLLHWNHGLKPMGR-----TASYDVEKRW 349

RESULT 15
US-09-991-163-171
; Sequence 171, Application US/09991163
; GENERAL INFORMATION:
; APPLICANT: Ashkenazi, Avi J.
; APPLICANT: Baker, Kevin P.
; APPLICANT: Botstein, David
; APPLICANT: Desnoyers, Luc
; APPLICANT: Eaton, Dan L.
; APPLICANT: Ferrara, Napoleone
; APPLICANT: Fong, Sherman
; APPLICANT: Gerber, Hanspeter
; APPLICANT: Gerritsen, Mary E.
; APPLICANT: Goddard, Audrey
; APPLICANT: Godowski, Paul J.
; APPLICANT: Grimaldi, J. Christopher
; APPLICANT: Gurney, Austin L.
; APPLICANT: Kljavin, Ivar J.
; APPLICANT: Napier, Mary A.
; APPLICANT: Pan, James
; APPLICANT: Paoni, Nicholas F.
; APPLICANT: Roy, Margaret Ann
; APPLICANT: Stewart, Timothy A.
; APPLICANT: Tumas, Daniel
; APPLICANT: Watanabe, Colin K.
; APPLICANT: Williams, P. Mickey
; APPLICANT: Wood, William I.
; APPLICANT: Zhang, Zemin
; TITLE OF INVENTION: Secreted and Transmembrane Polypeptides and Nucleic
; FILE REFERENCE: P2730PIC17
; CURRENT FILING DATE: 2001-11-14
; PRIOR APPLICATION NUMBER: 60/049787
; PRIOR FILING DATE: 1997-06-16
; PRIOR APPLICATION NUMBER: 60/062250
; PRIOR FILING DATE: 1997-10-17
; PRIOR APPLICATION NUMBER: 60/065186
; PRIOR FILING DATE: 1997-11-12
; PRIOR APPLICATION NUMBER: 60/065311
; PRIOR FILING DATE: 1997-11-13


```

; PRIOR APPLICATION NUMBER: 60/091544
; PRIOR FILING DATE: 1998-07-01
; PRIOR APPLICATION NUMBER: 60/091519
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091626
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091633
; PRIOR FILING DATE: 1998-07-02
; PRIOR APPLICATION NUMBER: 60/091978
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/091982
; PRIOR FILING DATE: 1998-07-07
; PRIOR APPLICATION NUMBER: 60/092182
; PRIOR FILING DATE: 1998-07-09

```

Query Match 7.2%; Score 118.5; DB 10; Length 371;

Best Local Similarity 20.4%; Pred. No. 0.00025;
Matches 48; Conservative 38; Mismatches 74; Indels 75; Gaps 9;

```

Qy 82 TYARLKLEVIADCKVLYLDTVLVRDGLKPLMDTDL----- 119
Db 151 TPARFYLPILVPSAKKAIYMDDDIYVGGDILALYNTALKPGHAAAFSEDCDSASTKVIR 210
Qy 120 --GG--NWGACIDLFVERQEGYKQKIGMADGEYFNAGVLLINLKKWRHDIFKMSCEW 175
Db 211 GAGNGYNYIG-----YLDYKKERIRKLSMKASTCSFNPGEVYANLLEWKQRQNTNQLEK 265
Qy 176 VBOYNDVWQYQODLNLGFLKGVGCYANSRFEFMTNYAFMANGFASRHT--DPLYLDR- 232
Db 266 MK-----LNVEEG-----YSRTLAGSITTPPLLIY--FYQOHSITDIPMMNVRH 307
Qy 233 -----TWTAMPVAVSHYCSGSAKPMHRDCTVWGAERFTELASLTVPPEW 277
Db 308 IGSSAGKRYSPQFVAAKLLHWNHGLKPPWGR-----TASYTDVWEKN 349

```

Search completed: December 2, 2002, 12:16:01
Job time : 7.13314 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 13.5805 Seconds
(without alignments)
2166.126 Million cell updates/sec

Title: US-10-007-267-4

Perfect score: 1657

Sequence: 1 MDIVFAADNVAAYLCVAAK.....KCMLOQRKKLSARFLRKIIY 306

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

283224

Total number of hits satisfying chosen parameters:

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	349	21.1	220	2 C64146	hypothetical prote
2	252	15.2	286	2 S16423	general stress pro
3	239	14.4	406	2 H95205	glycosyl transfera
4	233	14.1	398	2 G95205	glycosyl transfera
5	220	13.3	335	2 H91191	hypothetical prote
6	220	13.3	335	2 A86039	probable LPS biosy
7	210	12.7	337	2 AH0972	lipopolysaccharide
8	201.5	12.2	404	2 C95206	glycosyl transfera
9	191	11.5	339	2 C62982	lipopolysaccharide
10	190.5	11.5	346	2 T01514	lipopolysaccharide
11	190	11.5	517	2 A71707	lipopolysaccharide
12	186	11.2	361	2 C86157	hypothetical prote
13	184	11.1	572	2 T51525	hypothetical prote
14	180	10.9	338	2 S47847	lipopolysaccharide
15	179.5	10.8	814	2 A95206	glycosyl transfera
16	177	10.7	337	2 G86038	probable LPS biosy
17	177	10.7	337	2 F91191	hypothetical prote
18	176.5	10.4	696	2 D95206	hypothetical prote
19	171.5	10.4	357	2 T48055	hypothetical prote
20	170	10.3	20	2 G95922	probable glycosyl
21	168	10.1	336	2 A10972	lipopolysaccharide
22	162.5	9.8	341	2 T46153	hypothetical prote
23	159.5	9.6	394	2 A71963	probable lipopolys
24	157	9.5	372	2 G64539	lipopolysaccharide
25	157	9.5	390	2 F96723	hypothetical prote
26	156.5	9.4	658	2 T05655	hypothetical prote
27	153	9.2	589	2 F86202	hypothetical prote
28	152	9.2	393	2 T00647	glycosyl transfera
29	148.5	9.0	161	2 E97921	conserved hypother

30	144	8.7	537	2 T49162	hypothetical prote
31	139	8.4	673	2 T47905	hypothetical prote
32	136.5	8.2	582	2 F84807	hypothetical prote
33	134	8.1	335	2 G96607	probable galactino
34	133	8.0	361	2 T06009	hypothetical prote
35	132	8.0	328	2 T07610	W5176 protein - r1
36	131.5	7.9	363	2 A86267	hypothetical prote
37	130.5	7.9	377	2 E71968	probable lipopolys
38	127	7.7	373	2 F71884	probable lipopolys
39	127	7.7	921	2 F84593	hypothetical prote
40	124	7.5	61	2 A64005	hypothetical prote
41	123.5	7.5	429	2 T31001	hypothetical prote
42	122.5	7.4	344	2 A84912	probable galactino
43	117	7.1	528	2 F86226	hypothetical prote
44	115.5	7.0	528	2 D84903	hypothetical prote
45	112.5	6.8	334	2 H96629	hypothetical prote

ALIGNMENTS

RESULT 1

C64146 Hypothetical protein HI0259 - Haemophilus influenzae (strain Rd KW20)

C:Species: Haemophilus influenzae

C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999

C:Accession: C64146

R:Flaeschmann, R.D.; Adams, M.P.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage

, D.M.; Brandon, R.C.; Scott, J.; Shirley, R.; Liu, L.I.; Glodok, A.; Kelley, J.M.; Weidman

, Science 269, 496-512, 1995

A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente

A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.

A:Reference number: A64000; MUID:95350630; PMID:7542800

A:Accession: C64146

A:Status: nucleic acid sequence not shown; translation not shown

A:Molecule type: DNA

A:Residues: 1-220 <TIGR>

A:Cross-references: GB:I42023; TIGR:HI0259; GB:U32711; NID:g1573214; PIDN:AA021924.1;

Query Match 21.1%, Score 349, DB 2, Length 220;

Best Local Similarity 40.2%, Pred. No. 3.2e-23;

Matches 72; Conservative 39; Mismatches 62; Indels 6; Gaps 5;

Qy	1	MDIVFAADNVAAYLCVAAKSYEAHHPDTEIRFHYLDAGISENRAAVALRGSGNIR-	59
Db	39	MNIFSSDHYTAIPLAVSIFSLIKNTP-KTINFTLDKMKINOENK-TIINNLAAYSKV	96
Qy	60	-FIDVNPEDFAGFPLNIRHISITTYARLKGXYIADCCKVLYLDVLRDGLKPLMDTD	118
Db	97	FELPYCESDFQNFRTIDYISLATYARLNLTKYINIKAIYIDVDTLTNSLSLOELMND	156
Qy	119	LGGMVAGACIDLFVE-ROEGYKQKIGMADGEYFNAGVLLNLKKRRHDIFFKSCENV	176
Db	157	ITNYLIAACRDFIDVKNKAYKKTIGL-EGSYFNAGJILLNLNKKRENIFFOKSINNM	214

RESULT 2

S16423 general stress protein gspa - Bacillus subtilis

C:Species: Bacillus subtilis

C:Date: 06-Jan-1994 #sequence_revision 24-Feb-1995 #text_change 08-Dec-2000

C:Accession: S16423; S39667; P50148; A69638

R:Glaser, P.; Kunst, F.; Debarbouille, M.; Vertes, A.; Danchin, A.; Dedonder, R.

DNA Seq. 1, 251-261, 1991

A:Title: A gene encoding a tyrosine tRNA synthetase is located near sacs in Bacillus

A:Reference number: S16421; MUID:92216127; PMID:1806041

A:Accession: S16423

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 7-286 <GL2>

A:Cross-references: EMBL:X52480; NID:940236; PIDN:CAA36721.1; PID:940239

R:Glaser, P.; Kunst, F.; Arnaud, M.; Coudart, M.P.; Gonzales, W.; Hullo, M.F.; Ionesc

[illegible]

RESULT 5

hypothetical protein ECs4504 [imported] - *Escherichia coli* (strain O157:H7, substrain R
C:Species: *Escherichia coli*
C:Date: 18-Jul-2001 #sequence,revision 18-Jul-2001 #text_change 03-Aug-2001
C:Accession: J91191
R:Hayashi, T., Makino, K.; Ohnishi, M.; Kurokawa, K.; Ishii, K.; Yokoyama, K.; Han, C.G.
gasawara, N.; Yasunaga, T.; Kuhara, S.; Shiba, T.; Hattori, M.; Shinagawa, H.
DNA Res. 8, 11-22, 2001
A:Title: Complete genome sequence of enterohemorrhagic *Escherichia coli* O157:H7 and gene
A:Reference number: A99629; MUID:21156231; PMID:11258796
A:Accession: J91191
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-335 <HAV>
A:Cross-references: GB:BA000007; PTDN:BA37927.1; PID:g13363979; GSPDB:GN00154
A:Experimental source: strain O157:H7, substrain RMD 0509952
C:Genetics:
A:Gene: ECs4504
C:Superfamily: l1lipopolysaccharide glucosyltransferase I

Query Match	13.3%;	Score 220;	DB 2;	Length 335;
Best Local Similarity	26.9%;	Pred. No. 1.2e-11;		
Matches 73;	Conservative 45;	Mismatches 109;	Indels 44;	Gaps 13;

[illegible]

RESULT 6

Probable Irgs bishynthesis enzyme wai [Imported] - Escherichia coli (strain O157:H7, subsp. C)
C.Species: Escherichia coli
C.Date: 16-Feb-2001
C.Sequence_revision: 16-Feb-2001
C.#text_change: 14-Sep-2001
C.Accession: A86039
R.Perna, N.T.; Plunkett III, G.; Burland, V.; Mau, B.; Glasner, J.D.; Rose, D.J.; Mayhew, M.W.; Fink, L.L.; Grobeck, E.J.; Davis, N.W.; Lim, A.; Diala, E.; Potamousis, K.; Apodaca, Nature 409, 529-533, 2001
Title: Genome sequence of enterohemorrhagic Escherichia coli O157:H7.

A:Reference number: AB5480; MUID:21074935; PMID:11206551
A:Accession: AB6039
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-335 <STO>
A:Cross-references: GB:AE005174; NID:g12518362; PIDN:AAG58773.1; GSPDB:GN00145; UWGP
A:Experimental source: strain O157:H7, substrain EDL933
C:Genetics:
A:Gene: waal
C:Superfamily: 1lipopolysaccharide glucosyltransferase I

Query Match	13.3%	Score 220;	DB 2;	Length 335;
Best Local Similarity	26.9%	Pred. NO.1.2e-11;		
Matches 73;	Conservative 45;	Mismatches 109;	Indels 44;	Gaps 13;

[illegible]

RESULT 7

lipopolysaccharide 3-alpha-galactosyltransferase (EC 2.4.1.44) - *Salmonella enterica*
A:Note: This species has also been called *Salmonella typhi*
C:Date: 09-Nov-2001 #sequence_revision 09-Nov-2001 #text_change 27-Nov-2001
C:Accession: AH0972
R:Parkhill, J.; Dougan, G.; James, K.D.; Thomson, N.R.; Pickard, D.; Wain, J.; Church-
th, T.; Connerton, P.; Cronin, A.; Davis, P.; Davies, R.M.; Dowd, L.; White, N.; Farrar,
S.; Mole, S.; O'Garra, P.
Nature 413, 848-852, 2001
A:Authors: Parry, C.; Quail, M.; Rutherford, K.; Simmonds, M.; Skelton, J.; Stevens,
A:Title: Complete genome sequence of a multiple drug resistant *Salmonella enterica* sero-
A:Reference number: AB0502; PMID:11677608
A:Accession: AH0972
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-337 <PAR>
A:Cross-references: GB:AL513382; PIDN:CAD03276.1; PID:G16504687; GSPDB:GN00176
C:Genetics:
A:Gene: waat
C:Superfamily: lipopolysaccharide glucosyltransferase I
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match	12.7%;	Score 210;	DB 2;	Length 337;
Best Local Similarity	25.9%;	Pred. No. 9.5e-11;		
Matches	72;	Conservative 47;	Mismatches 111;	Indels 48;
				Gaps 13

QY 2 DVEAADONNAAVILCVAAKSYEAAPDDEIFREHYLD--AGISEENRAAVANLNGGGINR 59
Db 29 NAYGIDKNLEFGGCVSIAASYIARREKALNFHYFTDFEGEDQORFADLAKOYATIV 88
QY 60 FIDVNPEDPAGEPLNIRHISTTARLKGXYIAD-CDKVLYTDVYLVRDGLKPLWDTD 118
Db 89 YL-IDCERKSLP-STKMWTAYATYFRFTIADYFSDKDRYVLLDADJACKSIOELIDLN 146
QY 119 LGGNVAGACIDIEFVROEG-----YKQI-----GMADGEYYPNAGVLLINLKWRRHDI 168

Db 147 FAENIAAVV-----ABGELEMTKRSVSLATPOLVSG--YFNAGFTLIIINPLMTAENI 198
Qy 169 FKMSIEWEQKKDWMO--YODODILNGLFKGGVCYANSRFPNPTNYAFMANGFASRHTD 226
Db 199 SKKALEMKDEPVVQRITTHDODVLNLLVNAKARFVDKRFMTQ-----FSLN----- 245
Qy 227 PLYLDRTNTAMPVAVS---HYCGSAKPMHRDCTVWGA 260
Db 246 --YELKDSVINPVDAEYFVHIGPTKPMHS---WGA 277

RESULT 8
glycosyl transferase, family 8 SP1770 [imported] - Streptococcus pneumoniae (strain TIGR C95206
C:Species: Streptococcus pneumoniae
C:Date: 03-Aug-2001 #sequence_revision 03-Aug-2001 #text_change 03-Aug-2001
C:Accession: C95206
R:Rettelin, H.; Nelson, K.E.; Paulsen, I.T.; Eisen, J.A.; Read, T.D.; Peterson, S.; Heid
on, J.D.; Umayam, L.A.; White, O.; Salzberg, S.L.; Lewis, M.R.; Radune, D.; Holtzaple,
nson, T.; Hickey, E.K.; Holt, I.E.
Science 293, 498-506, 2001
A:Authors: Loftus, B.J.; Yang, F.; Smith, H.O.; Venter, J.C.; Dougherty, B.A.; Morrison,
A:Title: Complete Genome Sequence of a virulent isolate of Streptococcus pneumoniae.
A:Reference number: A95000; MUID:21357209; PMID:11463916
A:Accession: C95206
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-404 <KUN>
A:Cross-references: GB:AE005672; PIDN:AAK75844.1; PID:g14973266; GSPDB:GN00164; TIGR:SP4
A:Experimental source: strain TIGR4
C:Genetics:
A:Gene: SP1770

Query Match 12.2%; Score 201.5; DB 2: Length 404;
Best Local Similarity 26.8%; Pred. No. 6.7e-10;
Matches 74; Conservative 48; Mismatches 107; Indels 47; Gaps 12;

Qy 3 IYFAADDNAAVLCVAKSVEAHPDTEIRFHVLDAGISEENRAAVANLRGGG---NI 58
Db 6 IYFNADNDVLDLETAIKSI--CCYNNCLKFVENDDISAEFMMKNRKLTIOSEIVNV 63
Qy 59 RRTDVNPEDFAGPPLNIRIISTTTARLKLGEYIADCDDVLYLDVLYVRDLKPLMDTD 118
Db 64 KIVD--HYLKKRFHLPLKMLSYATFFRYPINPVEK-SRALYLDSDIITGSLDYLFDE 119
Qy 119 LGNNWVGACIDLFVEHQEGYKQKIGMADEY---YFNAGVLLINLKKWRHDIKMSCEM 175
Db 120 LGGYLAALAEVDSF-----GDVPSTNFGMLVNVDTWDEDAKSLLEL 164
Qy 176 VEQYKDVMOYODDILNGLFKGGVCYANSRFPNPTNYAFMANGFASRHTDPLY-----L 230
Db 165 TNGYHET-AYGQGLIMMF-----HDMRWKRLDRNFNFNMGMDSVAHLEGHKKMYEIS 216
Qy 231 DRTNTAMPVAVSHYCGSAKPMHRDCTVWGAERFTEL 266
Db 217 ELKNGDLP-SVIHYTG-VKPMW-----IISNNRREV 246

RESULT 9
C42982
lipopolysaccharide 3-alpha-galactosyltransferase (EC 2.4.1.44) - Escherichia coli (strat
N:Alterate names: lipopolysaccharide core assembly protein rfaI; UDP-D-galactose (gluc
C:Species: Escherichia coli
C:Date: 04-Mar-1993 #sequence_revision 18-Nov-1994 #text_change 01-Mar-2002
C:Accession: C42982; S47848; E65163
R:Pradel, E.; Parker, C.T.; Schaitman, C.A.
J. Bacteriol. 174, 4736-4745, 1992
A:Title: Structures of the rfaB, rfaI, rfaD, and rfaE genes of Escherichia coli K-12 and
A:Reference number: A42982; MUID:92325066; PMID:1624461
A:Contents: K-12
A:Accession: C42982
A:Status: preliminary

A:Molecule type: DNA
A:Residues: 1-339 <PRA>
A:Cross-references: GB:M80599; GB:M86935; NID:g146650; PIDN:AAA24086.1; PID:g146656
A:Note: sequence extracted from NCBI backbone (NCBIN:108180, NCBIPI:108185)
R:Plunkett, G.
submitted to the EMBL Data Library, March 1994
A:Reference number: S47666
A:Accession: S47848
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-339 <PU>
A:Cross-references: EMBL:U00039; NID:g466582; PIDN:AA818604.1; PID:g466765
R:Blattner, F.R.; Plunkett III, G.; Bloch, C.A.; Perna, N.T.; Burland, V.; Riley, M.;
.A.; Rose, D.J.; Mau, B.; Snao, Y.
Science 277, 1453-1462, 1997
A:Title: The complete genome sequence of Escherichia coli K-12.
A:Reference number: A64720; MUID:97426617; PMID:9278503
A:Accession: E65163
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-339 <BLAT>
A:Cross-references: GB:AE00440; GB:U00096; NID:g2367252; PIDN:AACT6651.1; PID:g17900
A:Experimental source: strain K-12, substrain MG1655
C:Genetics:
A:Gene: rfaI
A:Start codon: GTG
C:Superfamily: Lipopolysaccharide glucosyltransferase I
C:Keywords: glycosyltransferase; hexosyltransferase

Query Match 11.5%; Score 191; DB 2: Length 339;
Best Local Similarity 21.4%; Pred. No. 4.5e-09;
Matches 67; Conservative 60; Mismatches 142; Indels 44; Gaps 12;

Qy 1 MDIVFAADDNAAVLCVAKSVEAHPDTEIRFHVLDAGISEENRAAV-AANLRGGNIR 59
Db 29 LDIAVGTDKNFLFGGGSISILKYNEGRSLCFHIFTDFGDDRKYPDALALAYKTRIK 88
Qy 60 FIDVNPEDFAGPPLNIRIISTTTARLKLGEY-IADCCKVLYLDVLYVRDLKPLMDTD 118
Db 89 IYLVNGDRLSLP-STKMTWTAHYFRFYADFIKARPVLYLDADITICQGTIEPLNFS 147
Qy 119 LGNNWVGACIDLFVEHQEGYKQK-----IGMADGEYFNAGVLLINLKKWRHDIKMS 171
Db 148 PDDIVYAMV---TEGQADWMEKRAHSLGVAIGAG--YFNAGFTLIIINPLMTAENI 202
Qy 172 SCEWVEQYKDV--MOYODDILNGLFKGGVCYANSRFPNPTNYAFMANGFASRHTDPL 228
Db 203 ATAMLENEPIIKITTHPDODVLNMLADKLFADIKYNTQFSLNYQKESFINPYTNDTI 262
Qy 229 YLDRTNTAMPVAVSHYCGSAKPMHRDCTVWG-----AERFTELAGSLTTPPEWNGKLAV 283
Db 263 FT-----HYIGTPKPHD---WANDYPVQAEMKAKNA-----SPWKNTALL 301
Qy 284 PPTKCMLOFRKK 296
Db 302 KPNNSNQLRYSAK 314

RESULT 10
T01514
glycosyl transferase homolog T10M13.14 - Arabidopsis thaliana
C:Species: Arabidopsis thaliana (mouse-ear cress)
C:Date: 19-Feb-1999 #sequence_revision 19-Feb-1999 #text_change 24-Mar-1999
C:Accession: T01514
R:Johnson, A.F.; de la Bastide, M.; Lodhi, M.; Hoffman, J.; Hasegawa, A.; Gnoj, L.; G
Martensen, R.; McCombie, W.
submitted to the EMBL Data Library, May 1997
A:Description: The sequence of the Arabidopsis thaliana T10M13 BAC.
A:Reference number: Z14346
A:Accession: T01514
A:Status: translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residues: 1-346 <JOH>

A:Cross-references: EMBL:AF001308; NID:g2104523; PID:g2104536
 A:Experimental source: cultivar Columbia
 C:Genetics:
 A:Map position: 4S
 A:Note: T10M13.14

Query Match 11.5%; Score 190.5; DB 2; Length 346;
 Best Local Similarity 26.3%; Pred. No. 5.1e-09;
 Matches 77; Conservative 36; Mismatches 101; Indels 79; Gaps 15;

OY 3 IVFAADNYAAYLCVAASYV-EAAHPTEIRFHVLDAGISEENR-----AAVANIRG 54
 DB VAITLDVYELGSIANAIVSIIOHSVCPES-VFHFRI--AVSEETNLESLSVSPRRLK 123
 OY 55 GGNIRFIVNPEDFAGF-----PLNTRHISITTYARLKGFIADOC-DKVLVD 102
 DB 124 ---FNIFYFAPETVAGLSSSVYQALDLEOPLN-----YASVYADLDLEPCVKNVYILD 172
 OY 103 TDVLVRDGLKPLMDTDLGNWVGA---CIDLFVERQEGYKOKIGMADGEY-----YFNA 153
 DB 173 SDLVVVDVADIAKLMTSLGSRITGAPEYCHANPTKYFTGGFSESEKFGSTFRGRKPCYFNT 232
 OY 154 GVLINLKKRRHDIFFKSCSEVEQYKDYMOYODDILNGL-----FKGVCYANSPRN 207
 DB 233 GVMVIDLKKMRGGYTKRIEKMEIQRERRIYE---LGSPLPFLVFSGVHAPISRHMN 288
 OY 208 FMPTVFAMNGF-----ASRHTDPLYLDRNTAMPVAVSHYCGSAKPMHR 253
 DB 289 -----QHGLGGDNVNGSGRDLHP-----GVSGLHMSGSGKPMWR 323

RESULT 11

lipopolysaccharide 1,2-glucosyltransferase (rfau) RP476 - Rickettsia prowazekii
 C:Species: Rickettsia prowazekii
 C:Date: 21-Nov-1998 #sequence_revision 21-Nov-1998 #text_change 03-Nov-2000
 C:Accession: A11707
 R:Andersson, S.G.E.; Zomorodipour, A.; Andersson, J.O.; Slicheritz-Ponten, T.; Alsmark, U
 Nature 396, 133-140, 1998
 A:Title: The genome sequence of Rickettsia prowazekii and the origin of mitochondria.
 A:Reference number: A116530; MUID:99039499; PMID:9823893
 A:Accession: A11707
 A:Status: preliminary; nucleic acid sequence not shown; translation not shown
 A:Molecule type: DNA
 A:Residues: 1-517 <SANT>
 A:Cross-references: GB:AJ235271; GB:AJ235269; NID:g3868717; PIDN:CAAL4931.1; PID:g386103
 A:Experimental source: strain Madrid E
 C:Genetics:
 A:Gene: rfau; RP476

Query Match 11.5%; Score 190; DB 2; Length 517;
 Best Local Similarity 22.0%; Pred. No. 9.4e-09;
 Matches 67; Conservative 51; Mismatches 114; Indels 72; Gaps 10;

OY 1 MDIVEAADNTAAV-LCYAASVEAAHPTEIRFHVLDAGISEENRAVAANLRGGNIR 59
 DB 240 LDIALTINDKYAIIHAAVYIASLNSDLSFYRFHI---VWDSNDPISQESINKLSMQ 295
 OY 60 FIDVPEPFAFPPLNIRHISIT-----TYARLKGFIADOCVLYLDTVL 106
 DB 236 YIRDSIDETFFPDILNKLALYKEIKFSNWPISLVMRYLFYDKIFPNLDSILYLDADIV 355
 OY 107 VMDGLKPLMDTDLGNWVGA---IDLFV-----ERQEGYKOKIGMADGEYFAGVLLNL 160
 DB 356 VLRLDNLAKKIDM-NMYIAAGSLDRAITYCIHKVOECKRNVAS-----YKNSGIVFLNL 410
 OY 161 KWRRHDIFFKSCSEVEQYKDYMOYODDILNGLFKGVCYANSPRNFMPTVFAMNGF 220
 DB 411 KMRREKQYNNILLETILKNSKCFSPDODLNVAFONVLYPLSMKWNFC-----459
 OY 221 ASRHTDPLYLDRNTAMPVAVSHYCGSAKPM-----HROCT 256
 DB 460 -----TYFEHOSPYFSYFILHY-AEPKPMTEKEELMKTNHDKLDKITOYVWRYREIT 511

OY 257 VWGA 260
 DB 512 PWNS 515

RESULT 12

hypothetical protein T14P4.1 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 02-Mar-2001 #sequence_revision 02-Mar-2001 #text_change 31-Dec-2001
 C:Accession: C86157
 R:Theologis, A.; Ecker, J.R.; Palm, C.J.; Federspiel, N.A.; Kaul, S.; White, O.; Alon
 Chiu, C.W.; Chung, M.K.; Conn, L.; Conway, A.B.; Conway, A.R.; Cressy, T.H.; Dewar,
 anser, N.F.; Hughes, B.; Hultzar, L.
 Nature 408, 816-820, 2000
 A:Authors: Hunter, J.L.; Jenkins, J.; Johnson-Hopson, C.; Khan, S.; Khaykin, E.; Kim,
 C.A.; Li, J.H.; Li, Y.; Lin, X.; Liu, S.X.; Liu, Z.A.; Lueros, J.S.; Maili, R.; Marzia
 Rizzo, M.; Rooney, T.; Rowley, D.; Sakano, H.
 A:Authors: Salzberg, S.L.; Schwartz, J.R.; Shinn, P.; Southwick, A.M.; Sun, H.; Tallo
 ker, M.; Wu, D.; Yu, G.; Fraser, C.M.; Venter, J.C.; Davis, R.W.
 A:Title: Sequence and analysis of chromosome 1 of the plant Arabidopsis.
 A:Reference number: A86141; MUID:21016719; PMID:11130712
 A:Accession: C86157
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-361 <STO>
 A:Cross-references: GB:AF005172; NID:g9972380; PIDN:AGI0630.1; GSPDB:GN00141
 C:Genetics:
 A:Map position: 1

Query Match 11.2%; Score 186; DB 2; Length 361;
 Best Local Similarity 27.6%; Pred. No. 1.3e-08;
 Matches 72; Conservative 32; Mismatches 99; Indels 58; Gaps 11;

OY 33 FHVLDAGISEENRAVAANLRGGNIREIDVNPEDFAF-----PLNTRHISIT 81
 DB 114 FHFLVSSSQNLSELRSTFPLKTLKYFAPERVQSLSSVYQALDLEOPLN----- 166
 OY 82 TYARLKGFIADOCV-LYLDITDVLVADGLKPLMDTDLGNWVGA---CIDLFVERQEG 137
 DB 167 -YARNVYLDLLEPCVKRYIYDSDLVVVDVLYKLMKTGLGORTIGAPYCHANFTKYFTG 225
 OY 138 ---YKOKIGMADGE--YFNAGVLLINLKKRRHDIFFKSCSEVEQYKDYMOYODDIL 191
 DB 226 GPMSKRNRNGFFKGNPCYFNTGVAVIDLKKRORPFRFKRIEKMEIQRERRIYE---L 281
 OY 192 NGL-----FKGVCYANSPRNFMPTVFAMNGFASRHTDPLYLDRNTAMPVAVSHYC 245
 DB 282 GSLPFLFLVFAGHVAPISHRMNQHGLGGDNVNGSGRDLHSG-----PVSLLHMS 330
 OY 246 GSAKPMHR-----DC---TWV 258
 DB 331 GSGKPMRLRDLKRLPCPLDTLW 351

RESULT 13

hypothetical protein T20K14_80 - Arabidopsis thaliana
 C:Species: Arabidopsis thaliana (mouse-ear cross)
 C:Date: 18-Aug-2000 #sequence_revision 18-Aug-2000 #text_change 18-Aug-2000
 C:Accession: T51525
 R:Sato, S.; Nakamura, Y.; Kaneko, T.; Kato, T.; Asamizu, E.; Kotani, H.; Tabata, S.;
 submitted to the Protein Sequence Database, August 2000
 A:Reference number: Z25394
 A:Accession: T51525
 A:Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1-572 <SANT>
 A:Cross-references: EMBL:AL911143
 A:Experimental source: cultivar Columbia; BAC clone T20K14
 C:Genetics:
 A:Map position: 5

Tue Dec 3 09:23:11 2002

us-10-007-267-4.rpt

Page 7

Job time : 15.5805 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 7.44739 Seconds
(without alignments)
1704.189 million cell updates/sec

Title: US-10-007-267-4
Perfect score: 1657
Sequence: 1 MDIVFADNDYAAVLCVAAK.....KCMLORWKRLSARLEKRY 306

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	% Match	Query Length	ID	Description
1	536	32.3	330	1 Y258_HAEIN	P43974 bacillillus
2	252	15.2	286	1 GSPA_BACSU	P25148 bacillillus su
3	219	13.2	337	1 RPAI_SALTY	P19816 salmonella
4	191	11.5	339	1 RPAI_ECOLI	P27128 escherichia
5	180	10.9	338	1 RPAI_ECOLI	P27129 escherichia
6	173	10.4	336	1 RPAI_SALTY	P19817 salmonella
7	110.5	6.7	1548	1 UGGG_DROME	Q09332 drosophila
8	103.5	6.2	1448	1 UGGG_DROME	Q09140 schizosacch
9	94.5	5.7	393	1 MN10_YEAST	P50108 saccharomyc
10	93	5.6	1236	1 TOP2_LEICH	O61078 leishmania
11	89	5.4	505	1 SCRY_SALTY	P22340 salmonella
12	88	5.3	283	1 V635_MERJA	Q58052 methanococ
13	87	5.3	394	1 OM1B_CHLPR	P23421 chlamydia t
14	86.5	5.2	1202	1 DPMV_ASCIM	P22374 asacbolus t
15	85	5.1	576	1 GRK6_MOUSE	O70293 mus musculu
16	84	5.1	804	1 VP5_MTV	P13266 wound tumor
17	83.5	5.0	332	1 GRK6_RAT	P97111 rattus norv
18	83	5.0	576	1 LAGE_HUMAN	O95461 homo sapien
19	82.5	5.0	756	1 LAGE_HUMAN	O95461 homo sapien
20	82.5	5.0	756	1 LAGE_HUMAN	O95461 homo sapien
21	81.5	4.9	404	1 GIGC_MYCTU	Q91417 mus musculu
22	81.5	4.9	464	1 REM7_ECOLI	P37741 mycobacteri
23	81	4.9	187	1 PYRE_METAC	P58859 methanosarc
24	81	4.9	389	1 OM1N_CHLPR	Q07430 chlamydia p
25	81	4.9	389	1 OM1N_CHLPR	P27435 chlamydia p
26	81	4.9	1113	1 MYS_DICDI	P34109 dictyosteli
27	80.5	4.9	219	1 OSEB_ECO57	Q8363 escherichia
28	80.5	4.9	618	1 GIG1_YEAST	P36143 saccharomyc
29	80	4.8	900	1 GGAB_BACSU	P46918 bacillus su
30	79.5	4.8	295	1 YF67_ARCFU	O28705 archaeoglob
31	79.5	4.8	332	1 GLYG_MOUSE	Q97062 mus musculu
32	79.5	4.8	332	1 GLYG_RAT	O08730 rattus norv
33	79	4.8	333	1 OM1K_CHLPR	Q9xbf4 chlamydia p

34	79	4.8	437	1 PEPG_LACDL	P94869 lactobacilli
35	79	4.8	557	1 PGMU_EMENI	O9p931 emericella
36	79	4.8	1411	1 YK63_CABEL	P34342 caenorhabdi
37	78.5	4.7	1019	1 IDE_RAT	P35559 rattus norv
38	78	4.7	422	1 LAGE_AGRD	P29822 agrobacteri
39	78	4.7	1694	1 IGAO_HAEIN	P44969 haemophilus
40	78	4.7	1702	1 IGA2_HAEIN	P45384 haemophilus
41	77.5	4.7	99	1 RS24_PYRAB	O9uy20 pyrococcus
42	77.5	4.7	219	1 OSEB_SALTY	O8aeq3 salmonella
43	77.5	4.7	305	1 MIAA_THEWA	O9wy25 thermotoga
44	77.5	4.7	461	1 SYC_SALTY	O8z8p6 salmonella
45	77.5	4.7	461	1 SYC_SALTY	O8z8p6 salmonella

ALIGNMENTS

```

RESULT 1
ID Y258_HAEIN STANDARD: PRT: 330 AA.
AC P43974; P44597; P71347;
DT 01-NOV-1995 (Rel. 32, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DE 16-OCT-2001 (Rel. 40, Last annotation update)
DE Hypothetical protein HI0258.
GN HI0258.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=RD / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen B.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd.";
RL Science 269:496-512(1995).
RN [2]
RP REVISIONS.
RA White O., Clayton R.A., Kerlavage A.R., Fleischmann R.D.;
RL submitted (SRP-1996) to the EMBL/Genbank/DBJ databases.
CC -I- SIMILARITY: TO N.GONORRHOEAE LGIC AND WEAK. TO B.SUBTILIS GSPA.
CC LIPOPOLYSACCHARIDE 1,3-GALACTOSYLTRANSFERASE (RPA1) AND TO
CC LIPOPOLYSACCHARIDE 1,2-GLUCOSYLTRANSFERASE (RPAU).
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL: U32711; AAC21924.1; -.
CC TIGR: HI0258; -.
CC DR InterPro: IPR002495; GT_8.
CC DR Pfam: PF01501; Glyco-transf_8; 1.
CC KW Hypothetical protein: Complete proteome.
CC SEQUENCE 330 AA: 39073 MW: 65A6CA0207F24F9A CRC64:
Query Match 32.3%: Score 536; DB 1; Length 330;
Best Local Similarity 36.9%: Pred. No. 1.1e-40;
Matches 113; Conservative 58; Mismatches 109; Indels 26; Gaps 8;

```



```

RT Maal homologs."
RL J. Biol. Chem. 273:8849-8859(1998).
RN [2]
RC SEQUENCE FROM N.A.
RX STRAIN-LT2 / SCSCL12 / ATCC 700720;
RA MEDLINE-21534948; PubMed-11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Gireal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoenking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RT Nature 413:852-856(2001).
RN [3]
RP PRELIMINARY SEQUENCE OF 1-197 FROM N.A.
RA Flock J.-I.;
RT Submitted (JUL-1990) to the EMBL/Genbank/DBJ databases.
RN [4]
RP PRELIMINARY SEQUENCE OF 198-337 FROM N.A.
RC STRAIN-LT2;
RX MEDLINE-91045080; PubMed-2235496;
RA Carstenius P., Flock J.-I., Lindberg A.;
RT "Nucleotide sequence of rfaI and rfaJ genes encoding
RT lipopolysaccharide glycosyl transferases from Salmonella
RT typhimurium."
RX Nucleic Acids Res. 18:6128-6128(1990).
CC -1- FUNCTION: ADDS THE GALACTOSE(1) GROUP ON THE GLUCOSE(1) GROUP
CC OF LPS.
CC -1- CATALYTIC ACTIVITY: UDP-galactose + lipopolysaccharide = UDP + D-
CC galactosyl-lipopolysaccharide.
CC -1- PATHWAY: Lipopolysaccharide core biosynthesis.
CC -1- SIMILARITY: TO E. COLI RFAI AND TO RFAJ.
CC -1- CAUTION: REF 3 AND REF 4 SEQUENCES DIFFER FROM THAT SHOWN DUE TO
CC FRAMESHIFTS AND OTHER SEQUENCING ERRORS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AF026386; AAC16412.1; -
DR EMBL: AE008873; AAL22577.1; -
DR EMBL: X53847; CA37841.1; ALT_FRAME.
DR PIR: S12097; S12097.
DR StyGene: SG10337; rfaI.
DR InterPro: IPR002495; GT_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
KW Lipopolysaccharide biosynthesis; Glycosyltransferase; Transferase;
KW Complete proteome.
FT CONFLICT 125 125 R -> RE (IN REF. 1).
FT CONFLICT 210 210 E -> Q (IN REF. 1).
SQ SEQUENCE 337 AA; 38905 MW; 6DB1FA1531AE25F7 CRC64;

Query Match 13.2%; Score 219; DB 1; Length 337;
Best Local Similarity 26.7%; Pred. No. 2.9e-12;
Matches 74; Conservative 45; Mismatches 112; Indels 46; Gaps 12;

```

```

OY 170 KMSCENEQYKDVQ--YDODILNGLFRGVCYANRNFMTNYAFANGFASRHTPD 227
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 200 KRAIEMLKDPPEVYQRIHTHDDQVLNIFLVNKAKEFVJDKKFTQ-----FSLN----- 245
OY 228 LYIDRTNTAMPNPAVS-----HYCGSAPWHRDCTVWGA 260
| : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB 246 -YELKDSVINPDVAETVFEVHYIGTPKRWHS---MGA 277

RESULT 4
RFAI_ECOLI
ID RFAI_ECOLI STANDARD; PRT; 339 AA.
AC P27128;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-AUG-1992 (Rel. 23, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipopolysaccharide 1,3-galactosyltransferase (EC 2.4.1.44).
GN RFAI OR WAAI OR B3627.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-92325066; PubMed-1624461;
RA Pradel E., Parker C.T., Schaitman C.A.;
RT "Structures of the rfaB, rfaI, rfaJ, and rfaS genes of Escherichia
RT coli K-12 and their roles in assembly of the lipopolysaccharide
RT core."
RT J. Bacteriol. 174:4736-4745(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-94316500; PubMed-8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT "Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RT Nucleic Acids Res. 22:2576-2586(1994).
CC -1- CATALYTIC ACTIVITY: UDP-galactose + lipopolysaccharide = UDP + D-
CC galactosyl-lipopolysaccharide.
CC -1- PATHWAY: Lipopolysaccharide core biosynthesis.
CC -1- SIMILARITY: TO S. TYPHIMURIUM RFAI AND TO RFAJ.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: M80599; AAA24086.1; -
DR EMBL: U00039; AAB18604.1; -
DR EMBL: AE000440; AAC7651.1; -
DR PIR: C42982; C42982.
DR EcoGene: EG11352; rfaI.
DR InterPro: IPR002495; GT_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
KW Lipopolysaccharide biosynthesis; Glycosyltransferase; Transferase;
KW Complete proteome.
SQ SEQUENCE 339 AA; 39423 MW; 068CA689C1E0C1E0 CRC64;

Query Match 11.5%; Score 191; DB 1; Length 339;
Best Local Similarity 21.4%; Pred. No. 9.4e-10;
Matches 67; Conservative 60; Mismatches 142; Indels 44; Gaps 12;

```

```
Db 89 IYIINGDRILSLP-STKKNWTHAIFRFVIADYFINKAPKVLIDADIICOGTLEPLINFS 147
QY 119 LGSNWVACIDLEVEREGYKOK-----IGMADGEYFNAGVLLINLKKWRHDIFKM 171
Db 148 FPDQKVAWVY---TEGADWMEKRAHSLGVAIAKG--YNSGFLILINTQMMAQOVSAR 202
QY 172 SCWEVOYKDV--MOYODODILNGLFKGVGYCANSRPNF-MPTNYAFMANGFASRHTDPL 228
Db 203 AIALMEPEIKRITFDPOVLLNMLADKLIFADIKYNTQFSINYQKESFINPVTNDTI 262
QY 229 YLDRNTAMVAVASHYCSGSKPMWRDCTVG-----AERTELAGSLTTYPEERKGLAV 283
Db 263 FI-----HYIGPTKPMWD---WAMDYPVSOAFMEAKNA-----SPWKNTALL 301
QY 284 PPTKCMLOFRMRK 296
Db 302 KPNNNSQLRYSK 314

RESULT 5
REFL_ECOLI STANDARD; PRT; 338 AA.
ID REFL_ECOLI
AC P27129;
DT 01-AUG-1992 (Rel. 23, Created)
DT 01-OCT-1994 (Rel. 30, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipopolysaccharide 1,2-glucosyltransferase (EC 2.4.1.58).
GN RFAJ OR WAAJ OR B3626.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia
OX NCBI_TaxId=562;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-K12;
RX MEDLINE-92325066; PubMed-1624461;
RA Pradel E., Parker C.T., Schnaitman C.A.;
RT Structures of the rfaH, rfaI, rfaJ, and rfaS genes of Escherichia
RT coli K-12 and their roles in assembly of the lipopolysaccharide
RT core."
RL J. Bacteriol. 174:4736-4745(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-K12 / MG1655;
RX MEDLINE-94316500; PubMed-8041620;
RA Sofia H.J., Burland V., Daniels D.L., Plunkett G. III, Blattner F.R.;
RT Analysis of the Escherichia coli genome. V. DNA sequence of the
RT region from 76.0 to 81.5 minutes."
RL Nucleic Acids Res. 22:2576-2586(1994).
CC -1- CATALYTIC ACTIVITY: UDP-glucose + lipopolysaccharide = UDP + D-
CC glucosyl-lipopolysaccharide.
CC -1- PATHWAY: Lipopolysaccharide core biosynthesis.
CC -1- SIMILARITY: TO S.TYPHIMURIUM RFAJ AND TO RFAI.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC
CC EMBL; M80599; AAA24087.1; -
CC DR EMBL; U00039; AAB18603.1; -
CC DR EMBL; AE000440; AAC76650.1; -
CC DR PIR; D42982; D42982.
CC DR EcoGene; EGI1353; rfaJ.
CC DR InterPro; IPR002495; GT_8.
CC DR Pfam; PF01501; Glyco_transf_8; 1.
CC DR Lipopolysaccharide biosynthesis; Glycosyltransferase; Transferase;
CC complete proteome.
CC FT CONFLICT 310 338 RYKHLVQHNYISGIAGVCYLCKRYRK -> DNIIF
CC (IN REF. 1).
```

```
SQ SEQUENCE 338 AA: 39040 MW: 819428EAL3F1959A CRC64;
Query Match 10.9%; Score 180; DB 1; Length 338;
Best local Similarity 23.7%; Pred. No. 9, 1e-09;
Matches 64; Conservative 48; Mismatches 118; Indels 40; Gaps 9;
QY 1 MDIVFADNDVYAALCYAAKSVFAHPDEIRFHVLD-----AGISEENRAVA 49
Db 28 LNVAVGDVANYLDVGVSITSIVLNNRHINLDFYIADVYNDGFQKIAKLARQNL--- 84
QY 50 ANLRGGNIFIDVNPEDFAGFLN--IRHSITTYARKLGEIYADCKVLYLDTVL 106
Db 85 -----RITLYRINDKLOCLPCTQWSRAMEFLFAFQLG---LTDRLLYLDADV 134
QY 107 VRDGLKPLMDTDGGMWVACIDLEVEREGYKOKIGMADGE---YFENAGVLLINLKKW 163
Db 135 CKGDISGLHLGLNGA-VAAVVNDVPEMOE--KAVSLSPPELLGQYFNSGVYLLDKW 191
QY 164 RRHDIFKSCWEVOYKDVMOYODODILNGLFKGVGYCANSRPNFMTNYAFMANGFASR 223
Db 192 ADAKLFKALSIILMSKDNVYKPPDQVMANVLLKGMTLFLPREYN---TIYIKSELKDT 248
QY 224 HTDPLVLDRTNTAMPVAVSHYCSGSAKPMWR 253
Db 249 HQNYKKLITESTLL---IHYGTATKPMWR 274

RESULT 6
REFL_SALTY STANDARD; PRT; 336 AA.
ID REFL_SALTY
AC P19817; 068266;
DT 01-FEB-1991 (Rel. 17, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lipopolysaccharide 1,2-glucosyltransferase (EC 2.4.1.58).
GN RFAJ OR WAAJ OR STM3717.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OX NCBI_TaxId=602;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN-LT2;
RX MEDLINE-98204873; PubMed-9535865;
RA Heinrichs D.E., Monteiro M.A., Perry M.B., Whitfield C.;
RT "The assembly system for the lipopolysaccharide R2 core-type of
RT Escherichia coli is a hybrid of those found in Escherichia coli K-12
RT and Salmonella enterica. Structure and function of the R2 waak and
RT waal homologs."
RL J. Biol. Chem. 273:8849-8859(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-LT2 / SGGC1412 / ATCC 700720;
RX MEDLINE-21534948; PubMed-11677609;
RA McClelland M., Sanderson K.E., Spieth J., Clifton S.W., Latreille P.,
RA Courtney L., Porwollik S., Ali J., Dante M., Du F., Hou S., Layman D.,
RA Leonard S., Nguyen C., Scott K., Holmes A., Grewal N., Mulvaney E.,
RA Ryan E., Sun H., Florea L., Miller W., Stoneking T., Nhan M.,
RA Waterston R., Wilson R.K.;
RT "Complete genome sequence of Salmonella enterica serovar Typhimurium
RT LT2."
RL Nature 413:852-856(2001).
RN [3]
RP SEQUENCE OF 1-241 FROM N.A.
RC STRAIN-LT2;
RX MEDLINE-91045080; PubMed-2255496;
RA Carstenius P., Flock J.-I., Lindberg A.;
RT "Nucleotide sequence of rfaI and rfaJ genes encoding
RT lipopolysaccharide glycosyl transferases from Salmonella
RT typhimurium."
RL Nucleic Acids Res. 18:6128-6128(1990).
RN [4]
RP SEQUENCE OF 242-336 FROM N.A.
```



```

DT 01-AUG-1991 (Rel. 19, Created)
DT 01-AUG-1991 (Rel. 19, last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Sucrose porin precursor.
GN SCRY.
OS Salmonella typhimurium.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Salmonella.
OC NCBI_Taxid=602;
RN [1]
RP SEQUENCE FROM N.A., AND SEQUENCE OF 23-33.
RC STRAIN=6153-62;
RX MEDLINE=91100329; PubMed=1846143;
RA Hãrdesty C., Ferran C., Dirlenzo J.M.;
RT "Plasmid-mediated sucrose metabolism in Escherichia coli:
RT characterization of scry, the structural gene for a
RT phosphoenolpyruvate-dependent sucrose phosphotransferase system
RT outer membrane porin."
RL J. Bacteriol. 173:449-456(1991).
RN [2]
RP SEQUENCE FROM N.A.
RX MEDLINE=91312133; PubMed=1649946;
RA Schmid K., Ebner R., Jahreis K., Tengeler J.W., Tilgemyer F.;
RT "A sugar-specific porin, Scry, is involved in sucrose uptake in
RT enteric bacteria."
RL Mol. Microbiol. 5:941-950(1991).
RN [3]
RP X-RAY CRYSTALLOGRAPHY (2.4 ÅNGSTRÖMS).
RX MEDLINE=98100081; PubMed=9437428;
RA Forst D., Wette W., Wacker T., Diederichs K.;
RT "Structure of the sucrose-specific porin Scry from Salmonella
RT typhimurium and its complex with sucrose."
RL Nat. Struct. Biol. 5:37-46(1998).
CC -1- FUNCTION: Porin for sucrose uptake.
CC -1- SUBUNIT: HOMOTRIMER.
CC -1- SUBCELLULAR LOCATION: Integral membrane protein. Outer membrane.
CC -1- DOMAIN: THE C-TERMINUS HELPS TO ANCHOR THE PORIN TO THE OUTER
CC MEMBRANE.
CC -1- SIMILARITY: BELONGS TO THE LAMB FAMILY OF PORINS.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; M38416; AAA98417.1; -
DR EMBL; X57400; CAA40656.1; -
DR PIR; S15193; S15193.
DR PIR; A39127; A39127.
DR PDB; 1AOS; 10-JUN-98.
DR PDB; 1AOT; 18-MAR-98.
DR InterPro; IPR003192; Porin_Lamb.
DR Pfam; PF02264; Lamb; 1.
KW Transport; Sugar transport; Outer membrane; Transmembrane; Porin;
KW Signal; Plasmid; 3D-structure.
FT SIGNAL 1 22
FT CHAIN 23 505 SUCROSE PORIN.
SQ SEQUENCE 505 AA; 55467 MW; DD37733EC22A2135 CRC64;
Query Match 5.4%; Score 89; DB 1; Length 505;
Best Local Similarity 19.3%; Pred. No. 2.1;
Matches 50; Conservative 35; Mismatches 82; Indels 92; Gaps 10;
QY 13 AYCYAAKSVSEAHAPTEI-----RFHVLDAGISE-ENRAAVANLRGGNIRF----- 60
DB 8 AMLIALLTSASASHAOTDITIFARLNALEFKRQEAENRQOTABENRGAAEKKVQOLITAO 67
QY 61 -----IDVNPEDFAGFPLINRHSITTYTARLKLEGIADCDKVLVLDPT 103

```

```

DB 68 OQKONSSTOEVAQRTARLEKKADKSGFEFH-----GYAR-----SGVIMNDS 110
QY 104 DVLVRG--LKPLMDDDLGNWNGACIDLFVEROBYKQ-----KIGADEYFF 151
DB 111 GASTKSGAVTTPAGETGATGIRLGNQADYVERMLEKHOTLDNGATTRFVWVADQTSI 170
QY 152 NAGVLLINLKWRHRHIFKMSCEVQYKDY-----MOYQODIINGLFKGCVCYANSR 205
DB 171 N-----DWTASTSDLVNRQAFVELGNLPTFAGPRKSGSTLMAGKR 209
QY 206 F-----NEMPTNYAFMA 217
DB 210 FDRDNFDIHWISDVVELA 228
RESULT 12
ID Y635_METUA STANDARD; PRT; 283 AA.
AC 058052;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, last sequence update)
DT 16-OCT-2001 (Rel. 40, last annotation update)
DE Hypothetical protein MJ0635.
GN MJ0635.
OS Methanococcus jannaschlii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_Taxid=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=96337999; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overberg R., Kirkness E.F., Weinstock K.G., Fuhmann J.M., Glöck A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Uiterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Woese C.R., Venter J.C.;
RT "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschlii."
RL Science 273:1058-1073(1996).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL; U67511; AAB98633.1; -
DR DR TIGR; M0635; -
KW Hypothetical protein; Complete proteome.
SQ SEQUENCE 283 AA; 32502 MW; ICE52FC8457C5E20 CRC64;
Query Match 5.3%; Score 88; DB 1; Length 283;
Best Local Similarity 22.1%; Pred. No. 1.3;
Matches 50; Conservative 47; Mismatches 65; Indels 64; Gaps 14;
QY 10 NYAAVLCVAKSVSEAHAPTEIRFHVLD-----GISENRAAVAN-----LRGGNIRF 60
DB 43 DYNGEYVYVFNKNIEDSKYSEYRSKILNALRLGNANGKGTGIYINTKRMIFSGSYDRF 102
QY 61 -----IDVNP-EDFAGFPL-----NIRHSITTY-----ARLKIGEX 91
DB 103 VILIEGDYDFDKRNKLKEIGVNPVEYSGFKLYTRPNDDKIALFLYKMIITAGTQGYV 162
QY 92 IADCKVLYLIDTVLYRD-GLKPLMD---TDLGGNNGACT-DLFVEROBYKQKTGMAD 146
DB 163 --DCINVINQEMDSLKKNREVMETIYDRLPD-----ACVEVSGTYSPPWYKT---VAE 210

```


GRK6_MOUSE STANDARD: PRT: 576 AA.

ID GRK6_MOUSE 070294: 070294: 30-MAY-2000 (Rel. 39, Created)

AC 30-MAY-2000 (Rel. 39, Last sequence update)

DT 30-MAY-2000 (Rel. 41, Last annotation update)

DT 15-JUN-2002 (Rel. 41, Last annotation update)

DE G protein-coupled receptor kinase GRK6 (EC 2.7.1.1) (G-protein-coupled receptor kinase 6).

GN GRK6 OR GRK6.

OS Mus musculus (Mouse).

OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi; Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Mus.

OX NCBI_Taxid:10090;

RN [1]

RP SEQUENCE FROM N.A. (ISOFORMS GRK6A AND GRK6B).

RA Premont R.T., Lefkowitz R.J.,

RT "Genomic organization of the GRK5 and GRK6 genes: conserved structure among the GRK subfamily of G protein-coupled receptor kinases and evolution of the GRK gene family."

RT Submitted (JAN-1998) to the EMBL/Genbank/DBJ databases.

RL [2]

RP SEQUENCE FROM N.A. (ISOFORMS GRK6A AND GRK6B).

RC STRAIN:129/SvJ, and C57BL/6J; TISSUE:Thymus;

RA Moepers B., Vatter P., Frode R., Waechler F., Gierschik P.;

RT "Primary structure of murine G-protein-coupled receptor kinase 6 splice variants predict differential regulation by posttranslational modifications."

RT Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.

CC -1- FUNCTION: SPECIFICALLY PHOSPHORYLATES THE ACTIVATED FORMS OF G PROTEIN-COUPLED RECEPTORS (BY SIMILARITY).

CC -1- SUBCELLULAR LOCATION: Membrane-bound.

CC -1- ALTERNATIVE PRODUCTS: 2 ISOFORMS; GRK6A (SHOWN HERE) AND GRK6B; ARE PRODUCED BY ALTERNATIVE SPLICING.

CC -1- SIMILARITY: BELONGS TO THE SER/THR FAMILY OF PROTEIN KINASES. GPCR SUBFAMILY.

CC -1- SIMILARITY: CONTAINS 1 RGS DOMAIN.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@sib-sib.ch).

CC -----

DR EMBL: AF040747: AAC09268.1: -

DR EMBL: AF040748: AAC09269.1: -

DR EMBL: AF040749: AAC09270.1: -

DR EMBL: Y17967: CAA76975.1: -

DR EMBL: Y17967: CAA76976.1: -

DR EMBL: Y15798: CAA75789.1: -

DR EMBL: Y15799: CAA75790.1: -

DR HSSP: Q63450: 1A06.

DR MGD: MGI:1347078: Gprk6.

DR InterPro: IPR000719: Euk_pkinase.

DR InterPro: IPR000239: GPCR_kinase.

DR InterPro: IPR000961: Pkinase_C.

DR InterPro: IPR000342: Regl_Gprotein.

DR InterPro: IPR002290: Ser_Thr_pkinase.

DR Pfam: PF00069: pkinase; 1.

DR Pfam: PF00615: RGS; 1.

DR PRINTS: PR00717: GPCR_KINASE.

DR PRODOM: PD000001: Euk_pkinase; 1.

DR SMART: SM00315: RGS; 1.

DR SMART: SM00133: S_TK_X; 1.

DR SMART: SM00220: S_TK_C; 1.

DR PROSITE: PS00107: PROTEIN_KINASE_ATP; 1.

DR PROSITE: PS00111: PROTEIN_KINASE_DOM; 1.

DR PROSITE: PS00108: PROTEIN_KINASE_ST; 1.

DR PROSITE: PS00132: RGS; 1.

KW Transferase; Serine/threonine-protein kinase; ATP-binding; Phosphorylation; Lipoprotein; Palmitate; Alternative splicing.

FT DOMAIN 1 185 N-TERMINAL.

FT	DOMAIN	186	448	PROTEIN KINASE.
FT	DOMAIN	449	576	C-TERMINAL.
FT	DOMAIN	53	171	RGS.
FT	NP_BIND	192	200	ATP (BY SIMILARITY).
FT	BINDING	215	215	ATP (BY SIMILARITY).
FT	ACT_SITE	311	311	BY SIMILARITY.
FT	MOD_RES	484	484	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	MOD_RES	485	485	PHOSPHORYLATION (AUTO-) (BY SIMILARITY).
FT	LIPID	561	561	PALMITATE (BY SIMILARITY).
FT	LIPID	562	562	PALMITATE (BY SIMILARITY).
FT	LIPID	565	565	PALMITATE (BY SIMILARITY).
FT	VARSPLIC	560	576	DCCGNCSDSEELPTRL -> RIAGVTAATVKKSSPPASSP QAEAPTGGWR (IN ISOFORM GRK6B).
FT	SEQUENCE	576 AA:	65978 MW;	91EC2028CAFA91 CRC64;

Query Match 5.1%; Score 85; DB 1; Length 576;

Best Local Similarity 21.3%; Pred. No. 5.7; Indels 96; Gaps 18;

Matches 71; Conservative 46; Mismatches 121;

32 REFVLDAGISEENRAA--VAANLRGGGIRFIDVNPEDPAGFPLNIRHSITTYARKLG 89

244 RF-VLSLAYAETKDALCLVLTLMNGDLKF-----HIYHGQAGFPARAV 289

90 EYIAD-CDRVLYLMDVLYVRDGLKP---LMD-----IDLGNWVGACIDLEVERQEGY 138

290 FYAAEICCGLEDLHRRIRIYRDLKPENILLDHGHIRISDLG-----LAVHVPGGQTI 342

139 KQKIG-----MADGEYR-----AGVLLINL-----KKRRHDIFKMSC 173

343 KGVGVGYMADEVYNNERYTSPDMWALGCLYEMIAQSPFOORKKIKREEVERLYK 402

174 EWVEQYKDYMOYODDILNGLFK-----GSCYANSRFNPFPNVAFMANGF-- 220

403 EYAEETDIFSSQARSCLCSQLSKDPAERLGGRGGARREYKHEPLFKTNFRLGAGMLE 462

221 -----ASRHTDPLYLDRNTAM-----PVAVSHY-----CGSAK-PWHRDCTWGAER 262

463 PPKPDPQAIFYCQDVLDIQFSTVKGVDLEPTDQDEYQKFGATGSVSI PMQNEV--ETEC 520

263 FTFL--AGSLITYPE--EMRGLIAPPTKMLQR 292

521 FDELNVFGLDGVSVPDLDKMGQPTAPPKKGLLOR 554

Search completed: December 2, 2002, 11:56:27

Job time : 8.44739 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 26.2849 Seconds
(without alignments)
2398.732 Million cell updates/sec

Title: US-10-007-267-4

Perfect score: 1657

Sequence: 1 MDIVFAADDNYAAYLCVAAK.....KCMLORMKKLSARFLRKIV 306

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

SPTREMBL_21:*
1: sp_archaea:*
2: sp_bacteria:*
3: sp_fungi:*
4: sp_human:*
5: sp_invertebrate:*
6: sp_mammal:*
7: sp_mhc:*
8: sp_organelle:*
9: sp_phage:*
10: sp_plant:*
11: sp_rodent:*
12: sp_virus:*
13: sp_vertebrate:*
14: sp_unclassified:*
15: sp_virus:*
16: sp_bacteriap:*
17: sp_archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1657	100.0	306	2 050948	050948 neisseria g
2	1631.5	98.5	307	2 09REX5	09REX5 neisseria g
3	1551.5	93.6	311	2 P96945	P96945 neisseria m
4	1550.5	93.6	311	2 Q93EK7	Q93EK7 neisseria m
5	572	34.5	302	2 09L6B2	09L6B2 pasteurella
6	572	34.5	302	16 09CLSO	09CLSO pasteurella
7	284	17.1	57	2 09REX6	09REX6 neisseria g
8	278	16.8	60	2 09RGNO	09RGNO neisseria g
9	269.5	16.3	269	2 09L7A2	09L7A2 haemophilus
10	257	15.5	314	2 09AHB5	09AHB5 streptococc
11	255	15.4	630	2 048480	048480 klebsiella
12	243	14.7	314	2 08VU29	08VU29 streptococc
13	239	14.4	406	16 097P76	097P76 streptococc
14	233	14.1	398	16 097P77	097P77 streptococc
15	230.5	13.9	290	2 068207	068207 escherichia
16	220	13.3	335	2 09ZIT4	09ZIT4 escherichia

17	220	13.3	335	16 08XDC0	08XDC0 escherichia
18	219.5	13.2	273	2 093RH6	093RH6 streptococc
19	219.5	13.2	635	2 048484	048484 klebsiella
20	218.5	13.2	342	2 09ZIS1	09ZIS1 escherichia
21	217.5	13.1	338	2 09R9D1	09R9D1 escherichia
22	216.5	13.1	338	2 09ZIS5	09ZIS5 escherichia
23	212.5	12.8	341	2 09ZIS8	09ZIS8 escherichia
24	210	12.7	337	16 08Z2G3	08Z2G3 salmonella
25	201.5	12.2	404	16 097P73	097P73 streptococc
26	197.5	11.9	309	2 052832	052832 rhizobium 1
27	194	11.7	331	2 09R9D2	09R9D2 escherichia
28	193	11.6	331	2 09ZIS6	09ZIS6 escherichia
29	190.5	11.5	346	10 004253	004253 arbidopsis
30	190	11.5	517	16 09ZD68	09ZD68 rickettsia
31	189	11.4	339	2 068205	068205 escherichia
32	187	11.3	338	2 068204	068204 escherichia
33	186	11.2	361	10 09FWY9	09FWY9 arbidopsis
34	184	11.1	572	10 09LF35	09LF35 arbidopsis
35	183	11.0	510	10 09MAB8	09MAB8 arbidopsis
36	179.5	10.8	814	16 097P75	097P75 streptococc
37	177	10.7	329	2 09ZNA6	09ZNA6 erythrobact
38	177	10.7	337	16 08XDC2	08XDC2 escherichia
39	176.5	10.7	696	16 087P72	087P72 streptococc
40	171.5	10.4	357	10 09LZJ9	09LZJ9 arbidopsis
41	171.5	10.4	361	10 08VYF4	08VYF4 arbidopsis
42	171	10.3	682	2 09AEU2	09AEU2 streptococc
43	170	10.3	337	16 09ZV02	09ZV02 rhizobium m
44	169.5	10.2	351	10 09M8J2	09M8J2 arbidopsis
45	168	10.1	336	16 08Z2G2	08Z2G2 salmonella

ALIGNMENTS

RESULT 1
Q50948 PRELIMINARY: PRT; 306 AA.
ID Q50948
AC 050948;
DT 01-NOV-1996 (TREMBLrel. 01, Created)
DT 01-NOV-1996 (TREMBLrel. 01, Last sequence update)
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)
DE Glycosyl transferase.
LN LGTC.
OS Neisseria gonorrhoeae.
OC Bacteria: Proteobacteria: beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=P62;
RX MEDLINE=95053752; PubMed=7964493;
RA Gotschlich E.C.;
RT "Genetic locus for the biosynthesis of the variable portion of
RT Neisseria gonorrhoeae lipooligosaccharide.";
RT J. Exp. Med. 180:2181-2190(1994).
DR EMBL; U14554; AAA68011.1; .
DR InterPro; IPR002495; GT_8.
DR Pfam; PF01501; Glyco_transf_8; 1.
KW Transferase.
SQ SEQUENCE 306 AA; 34857 MW; 17BB5208FD88AACB CRC64;
QY
QY Query Match 100.0%; Score 1657; DB 2; Length 306;
QY Best Local Similarity 100.0%; Pred. No. 9.4e-142;
QY Matches 306; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 1 MDIVFAADDNYAAYLCVAAKSYEAHPDTEIRRHVIDAGISEENRAVAANLRGGCNIRF 60
QY |||||||
QY 1 MDIVFAADDNYAAYLCVAAKSYEAHPDTEIRRHVIDAGISEENRAVAANLRGGCNIRF 60
QY |||||||
QY 1 IDVPDPAAGFPLNTHISTTAYARKLGEYTDCKVLYLDVDVLRGGRKLPMDTG 120
QY |||||||
QY 61 IDVPDPAAGFPLNTHISTTAYARKLGEYTDCKVLYLDVDVLRGGRKLPMDTG 120
QY |||||||
QY 61 IDVPDPAAGFPLNTHISTTAYARKLGEYTDCKVLYLDVDVLRGGRKLPMDTG 120
QY |||||||
QY 121 GNVVGACIDLEVEROBGYKOKIGMADGEYFFNAGVLLINLKKWRHRDIFKSCWEYQYK 180

```

|||||
Db 121 GNMWVACIDLFVEROEGYKQKIGMADGEYFNAGVLLINKKRRHDIKMSCEWEQYK 180
181 DVWQYODDOLNLGFKGVCYANSRFPNPTNAPFANGASRHTDPLVDRNTAMPVA 240
181 DVWQYODDOLNLGFKGVCYANSRFPNPTNAPFANGASRHTDPLVDRNTAMPVA 240
241 VSHYCGSAPMWHRDCTVWGAERFTELASLTTPVEEMRGKLAVPPTKCMLORRKKLSAR 300
241 VSHYCGSAPMWHRDCTVWGAERFTELASLTTPVEEMRGKLAVPPTKCMLORRKKLSAR 300
301 FLRKITY 306
301 FLRKITY 306
301 FLRKITY 306

RESULT 2
Q9REX5 PRELIMINARY; PRT; 307 AA.
ID Q9REX5;
AC Q9REX5;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE LgtC.
GN
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=485;
[1]
RP SEQUENCE FROM N.A.
RA Balthazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
RT "Mutations in the lgt operon influence serum-resistance in gonococci.";
RT gonococci";
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL; AF208062; AAF20991.1; -.
DR InterPro; IPR002495; GT_8.
DR Pfam; PF01501; Glyco_transf_8; 1.
SQ SEQUENCE 307 AA; 34903 MW; 02FB7015CB1CE635 CRC64;

```

```

Query Match 98.5%; Score 1631.5; DB 2; Length 307;
Best Local Similarity 99.0%; Pred. No. 1.9e-139;
Matches 304; Conservative 1; Mismatches 1; Indels 1; Gaps 1;

```

```

QY 1 MDIVFAADDNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAVAANLR-GGGNIR 59
1 MDIVFAADDNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAVAANLRGGGGNIR 60
60 FIDVNEDEAGFPLNIRHISITTYARLKIGEYIADCCKVLYLDTDVLVRDGLKPLMDTDL 119
61 FIDVNEDEAGFPLNIRHISITTYARLKIGEYIADCCKVLYLDTDVLVRDGLKPLMDTDL 120
120 GGNWVACIDLFVEROEGYKQKIGMADGEYFNAGVLLINKKRRHDIKMSCEWEQY 179
121 GGNWVACIDLFVEROEGYKQKIGMADGEYFNAGVLLINKKRRHDIKMSCEWEQY 180
180 KDVMOYODDOLNLGFKGVCYANSRFPNPTNAPFANGASRHTDPLVDRNTAMPV 239
181 KDVMOYODDOLNLGFKGVCYANSRFPNPTNAPFANGASRHTDPLVDRNTAMPV 240
240 AVSHYCGSAPMWHRDCTVWGAERFTELASLTTPVEEMRGKLAVPPTKCMLORRKKLSA 299
241 AVSHYCGSAPMWHRDCTVWGAERFTELASLTTPVEEMRGKLAVPPTKCMLORRKKLSA 300
300 FLRKITY 306
301 FLRKITY 307

```

```

RESULT 3
P96945 PRELIMINARY; PRT; 311 AA.
ID P96945;
AC P96945;

```

```

DT 01-MAY-1997 (TReMBLrel. 03, Created)
DT 01-MAY-1997 (TReMBLrel. 03, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Glycosyl transferase.
GN
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=487;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=20055626; PubMed=10589709;
RA Jennings M.P., Srikhanta Y.N., Moxon E.R., Kramer M., Poolman J.T.,
RA Kuipers B., van der Ley P.;
RT "The genetic basis of the phase variation repertoire of lipopolysaccharide immunotypes in Neisseria meningitidis.";
RT Microbiology 145:3013-3021(1999).
DR EMBL; U65788; AAB48385.1; -.
DR InterPro; IPR002495; GT_8.
DR Pfam; PF01501; Glyco_transf_8; 1.
DR Transferase.
SQ SEQUENCE 311 AA; 35743 MW; A33D91283857750B CRC64;

```

```

Query Match 93.6%; Score 1551.5; DB 2; Length 311;
Best Local Similarity 93.9%; Pred. No. 3.4e-132;
Matches 292; Conservative 2; Mismatches 12; Indels 5; Gaps 2;

```

```

QY 1 MDIVFAADDNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAVAANLR-GGGNIR 59
1 MDIVFAADDNYAAYLCVAAKSVAAHPDTEIRPHVLDAGISEENRAVAANLRGGGGNIR 60
60 FIDVNEDEAGFPLNIRHISITTYARLKIGEYIADCCKVLYLDTDVLVRDGLKPLMDTDL 119
61 FIDVNEDEAGFPLNIRHISITTYARLKIGEYIADCCKVLYLDTDVLVRDGLKPLMDTDL 120
120 GGNWVACIDLFVEROEGYKQKIGMADGEYFNAGVLLINKKRRHDIKMSCEWEQY 179
121 GNMWVACIDLFVEROEGYKQKIGMADGEYFNAGVLLINKKRRHDIKMSCEWEQY 180
180 KDVMOYODDOLNLGFKGVCYANSRFPNPTNAPFANGASRHTDPLVDRNTAMPV 239
181 KDVMOYODDOLNLGFKGVCYANSRFPNPTNAPFANGASRHTDPLVDRNTAMPV 240
240 AVSHYCGSAPMWHRDCTVWGAERFTELASLTTPVEEMRGKLAVP-PTKCMLORRKK 295
241 AVSHYCGSAPMWHRDCTVWGAERFTELASLTTPVEEMRGKLAVPPTKCMLORRKK 300
296 KLSARFLRKITY 306
301 KLSARFLRKITY 311

```

```

RESULT 4
Q93EK7 PRELIMINARY; PRT; 311 AA.
ID Q93EK7;
AC Q93EK7;
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE LgtC.
GN
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
NCBI_TaxID=487;
[1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21467954; PubMed=11583844;
RA Zhu P., Klutch M.J., Tsai C.-M.;
RT "Genetic Analysis Of Conservative and Variation of Lipooligosaccharide Expression in Two L8-Immunotype Strains of Neisseria meningitidis.";
RT FEMS Microbiol. Lett. 203:173-177(2001).
DR EMBL; AF355193; AAL12839.1; -.

```

DR InterPro: IPR002495; GT_8.
 Pfam: PF01501; Glyco_transf_8; 1.
 SO SEQUENCE 311 AA; 35738 MW; A6EDE3F85AC5B5D CRC64;

Query Match 93.6%; Score 1550.5; DB 2; Length 311;
 Best Local Similarity 93.9%; Pred. No. 4.2e-132;
 Matches 292; Conservative 2; Mismatches 12; Indels 5; Gaps 2;

QY 1 MDIVFADNDNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVANLR-GGNGIR 59
 DB 1 MDIVFADNDNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVANLRGGGNGIR 60
 QY 60 FIDVNPEDFAGPEPLNIRHISITTYARLKGEYIADCDKVLVLDTVLVDRGKPLMDTDL 119
 DB 61 FIDVNPEDFAGPEPLNIRHISITTYARLKGEYIADCDKVLVLDTVLVDRSLPLMDTDL 120
 QY 120 GGNWVGACIDLFVEROEGKOKIGMADGEYFNFAGVLLINLKKWRHDFKMSCEVVEQ 179
 DB 121 GGNWVGACIDLFVEROEGKOKIGMADGEYFNFAGVLLINLKKWRHDFKMSCEVVEQ 180
 QY 180 KDVMOYODDILNGLFKGVCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMPY 239
 DB 181 KDVMOYODDILNGLFKGVCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAMPY 240
 QY 240 AVSHICGSAKPMHRDCTVGAERFTELASLTTPPEMRGKLAVP---PTKCMLOMRK 295
 DB 241 AVSHYCGPAKPMHRDCTVGAERFTELASLTTPPEMRGKLAVPHRMSTKMLQWR 300
 QY 296 KLSARFLRKTY 306
 DB 301 KLSARFLRKTY 311

RESULT 5
 Q9L6B2 PRELIMINARY; PRT; 302 AA.

AC Q9L6B2; PRELIMINARY; PRT; 302 AA.
 DT 01-OCT-2000 (TREMBlrel. 15, Created)
 DT 01-OCT-2000 (TREMBlrel. 15, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Putative glycosyl transferase.
 GN LGTC.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OC NCBI_Taxid=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Fuller T.E., Kennedy M.J., Lowery D.E.;
 RT Identification of Pasteurella multocida virulence genes in a
 RT septicemic mouse model using signature-tagged mutagenesis";
 RL Submitted (FEB-2000) to the EMBL/Genbank/DBJ databases.
 DR EMBL: AE237927; AAF68413.1; -
 DR InterPro: IPR002495; GT_8.
 DR Pfam: PF01501; Glyco_transf_8; 1.
 DR Transferase.
 SQ SEQUENCE 302 AA; 35688 MW; C987B83F77FB158 CRC64;

Query Match 34.5%; Score 572; DB 2; Length 302;
 Best Local Similarity 38.9%; Pred. No. 1.4e-43;
 Matches 121; Conservative 61; Mismatches 115; Indels 14; Gaps 7;

QY 1 MDIVFADNDNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVANLR-GGNGIR 59
 DB 1 MDIVFADNDNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVANLRGGGNGIR 60
 QY 60 FIDVNPEDFAGPEPLNIRHISITTYARLKGEYIADCDKVLVLDTVLVDRGKPLMDTD 118
 DB 60 FIDVNPEDFAGPEPLNIRHISITTYARLKGEYIADCDKVLVLDTVLVDRGKPLMDTD 119
 QY 119 LGSNWWGACIDLFVEROEG-KYKIGMADGEYFNFAGVLLINLKKWRHDFKMSCEVVE 177
 DB 120 VNNFLTAACDPSFLENKSEKKSISMSDEKYYFNAGVMLFNLDEWRKMDVFSRALDLA 179

QY 178 QYKDVMOYODDILNGLFKGVCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAM 237
 DB 180 MYPNOMIYODDILNGLFKGVCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAM 237
 QY 238 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEMRG--KLAVPPTKMLQWRK 295
 DB 238 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEMRG--KLAVPPTKMLQWRK 296
 QY 296 KLSARFLRKTY 306
 DB 292 RIRYKFRYQY 302

RESULT 6
 Q9CL50 PRELIMINARY; PRT; 302 AA.

AC Q9CL50; PRELIMINARY; PRT; 302 AA.
 DT 01-JUN-2001 (TREMBlrel. 17, Created)
 DT 01-JUN-2001 (TREMBlrel. 17, Last sequence update)
 DT 01-MAR-2002 (TREMBlrel. 20, Last annotation update)
 DE Hypothetical protein PM139.
 GN PM139.
 OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OC NCBI_Taxid=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Whittam T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida pm70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL: AE006155; AKK03223.1; -
 DR InterPro: IPR002495; GT_8.
 DR Pfam: PF01501; Glyco_transf_8; 1.
 DR Hypothetical protein: Complete proteome.
 SQ SEQUENCE 302 AA; 35716 MW; 6C7264ED8F033C00 CRC64;

Query Match 34.5%; Score 572; DB 16; Length 302;
 Best Local Similarity 38.9%; Pred. No. 1.4e-43;
 Matches 121; Conservative 61; Mismatches 115; Indels 14; Gaps 7;

QY 1 MDIVFADNDNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVANLR-GGNGIR 59
 DB 1 MDIVFADNDNYAAYLCVAAKSVFAHPDTEIRFHVLDAGISEENRAAVANLRGGGNGIR 60
 QY 60 FIDVNPEDFAGPEPLNIRHISITTYARLKGEYIADCDKVLVLDTVLVDRGKPLMDTD 118
 DB 60 FIDVNPEDFAGPEPLNIRHISITTYARLKGEYIADCDKVLVLDTVLVDRGKPLMDTD 119
 QY 119 LGSNWWGACIDLFVEROEG-KYKIGMADGEYFNFAGVLLINLKKWRHDFKMSCEVVE 177
 DB 120 VNNFLTAACDPSFLENKSEKKSISMSDEKYYFNAGVMLFNLDEWRKMDVFSRALDLA 179
 QY 178 QYKDVMOYODDILNGLFKGVCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAM 237
 DB 180 MYPNOMIYODDILNGLFKGVCYANSRNFNPTNYAFMANGFASRHDPYLDRTNTAM 237
 QY 238 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEMRG--KLAVPPTKMLQWRK 295
 DB 238 PVAVSHYCGSAKPMHRDCTVGAERFTELASLTTPPEMRG--KLAVPPTKMLQWRK 296
 QY 296 KLSARFLRKTY 306
 DB 292 RIRYKFRYQY 302

RESULT 7
 Q9REX6 PRELIMINARY; PRT; 57 AA.
 AC Q9REX6; PRELIMINARY; PRT; 57 AA.

DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE LGTC.
GN LGTC.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=PA19;
RA Baltazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
RT "Mutations in the lgt operon influence serum-resistance in gonococci."
RL Submitted (NOV-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF208061; AAF20990.1; -
DR InterPro: IPR002495; GT_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
SQ SEQUENCE 57 AA; 5934 MW; 392247E1A7C1135D CRC64;

Query Match 17.1%; Score 284; DB 2; Length 57;
Best Local Similarity 100.0%; Pred. No. 1.6e-18;
Matches 56; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVFAADDNYAAYLCVAAKSVEAHDPTEIRFHVLDAGISEENRAVAANLRGG 56
Db 1 MDIVFAADDNYAAYLCVAAKSVEAHDPTEIRFHVLDAGISEENRAVAANLRGG 56

RESULT 8
09RGNO PRELIMINARY; PRT; 60 AA.

AC 09RGNO;
DT 01-MAY-2000 (TReMBLrel. 13, Created)
DT 01-MAY-2000 (TReMBLrel. 13, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE Glycosyltransferase.
GN LGTC.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1291;
RX MEDLINE=20305049; PubMed-10844691;
RA Harvey H.A., Porat N., Campbell C.A., Jennings M., Gibson B.W.,
RT Phillips N.J., Apicella M.A., Balke M.S.;
RT "Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein receptor on human sperm."
RL MOL. Microbiol. 36:1059-1070(2000).
DR EMBL: AF121135; AAF14361.1; -
DR InterPro: IPR002495; GT_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
KM Transferase.
SQ SEQUENCE 60 AA; 6304 MW; D28468118F7E1A7 CRC64;

Query Match 16.8%; Score 278; DB 2; Length 60;
Best Local Similarity 100.0%; Pred. No. 5.9e-18;
Matches 55; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MDIVFAADDNYAAYLCVAAKSVEAHDPTEIRFHVLDAGISEENRAVAANLRGG 55
Db 1 MDIVFAADDNYAAYLCVAAKSVEAHDPTEIRFHVLDAGISEENRAVAANLRGG 55

RESULT 9
09L7A2 PRELIMINARY; PRT; 269 AA.

AC 09L7A2;
DT 01-OCT-2000 (TReMBLrel. 15, Created)
DT 01-OCT-2000 (TReMBLrel. 15, last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, last annotation update)
DE Putative glycosyl transferase.

OS Haemophilus ducreyi.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=730;
RN [1]
RP SEQUENCE FROM N.A.
RA Wang J., Hanson E., Munson R.S. Jr.;
RT "Putative glycosyl transferase."
RL Submitted (JAN-2000) to the EMBL/GenBank/DBJ databases.
DR EMBL: AF224467; AAF28363.1; -
DR InterPro: IPR002495; GT_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
KM Transferase.
SQ SEQUENCE 269 AA; 32334 MW; 1F2E25A243AAC64 CRC64;

Query Match 16.3%; Score 269.5; DB 2; Length 269;
Best Local Similarity 30.8%; Pred. No. 2.9e-16;
Matches 78; Conservative 43; Mismatches 111; Indels 21; Gaps 7;

QY 1 MDIVFAADDNYAAYLCVAAKSVEAHDPTEIRFHVLDAGISEENRAVAANLRG-GGNIR 59
Db 8 MNIVLAANOSYSEYILTTISITL--HNKHIREYLLNRDYPTEFDILNKKLRKLNSEIT 65
QY 60 FIDVNPEDFAFPPLNIRHISITTYARIKGEYIADCDKVLVLDYLVYRDGLKPLMDTDL 119
Db 66 DIKVTNDTIKFKYTSYHISSDTEFFRYISDFI-EDQKVIYLDADIVYNGSLTELYQTDI 124
QY 120 GGNWVGACIDLFVEROGEYKQIGMADGEYFNAGVILLNKKWRHDIKMSCENEYQ 179
Db 125 SNYFLAAVKDIISEK-----IYNNHIFNAGMLLNKKWRHNTTQFLSISEKY 175
QY 180 KDVMQYODDIIINGLFKGGVYCANSEFNFV-PTNYAFMANGFASRHTDPLYIDRTYAMP 238
Db 176 INSLPADQSIILNIFKDKMLKRNRCYNIIGIDYLFKFG-----KTYILEDIGETIP 229
QY 239 VAVSHYCGSAPW 251
Db 230 LIH-HYNTKAPW 241

RESULT 10
09AHB5 PRELIMINARY; PRT; 314 AA.

AC 09AHB5;
DT 01-JUN-2001 (TReMBLrel. 17, Created)
DT 01-JUN-2001 (TReMBLrel. 17, last sequence update)
DT 01-JUN-2002 (TReMBLrel. 21, last annotation update)
DE wcin (CpsP) (Putative galactosyl transferase) (Putative galactosyl transferase Cps6ap).
GN wcin OR CpsP OR Cps6ap.
OS Streptococcus pneumoniae.
OC Bacteria; Firmicutes; Bacillus/Clostridium group; Lactobacillales;
OC Streptococcaceae; Streptococcus.
OX NCBI_TaxID=1313;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21116929; PubMed-11179285;
RA Jiang S.M., Wang L., Reeves P.R.;
RT "Molecular characterization of Streptococcus pneumoniae type 4, 6B, 8, and 18C capsular polysaccharide gene clusters."
RL Infect. Immun. 69:1244-1255(2001).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=PN93/1413;
RA Griffiths D.B., Hall L.M.C.;
RT "The Cps locus of Streptococcus pneumoniae serotype 6B, genes between cpsB and cpsL."
RL Submitted (AUG-2000) to the EMBL/GenBank/DBJ databases.
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=SSISF 64/1.
RA Griffiths D.B., Hall L.M.C.;
RT "The Capsular Polysaccharide Locus of Streptococcus pneumoniae


```

OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F632;
RX MEDLINE=98204873; PubMed=9535865;
RA Heinrichs D.E., Monteiro M.A., Perry M.B., Whitfield C.;
RT "The assembly system for the lipopolysaccharide R2 core-type of
RT Escherichia coli is a hybrid of those found in Escherichia coli K-12
RT and Salmonella enterica. Structure and function of the R2 Waak and
RT Maal homologs."
RL J. Biol. Chem. 273:8849-8859(1998).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=F632;
RX MEDLINE=99009352; PubMed=9791168;
RA Heinrichs D.E., Yethon J.A., Whitfield C.;
RT "Molecular basis for structural diversity in the core regions of the
RT lipopolysaccharides of Escherichia coli and salmonella enterica."
RL Mol. Microbiol. 30:221-232(1998).
DR EMBL: AF019375; AAC69655.1; -.
DR InterPro: IPR002495; GP_8.
DR Pfam: PF01501; Glyco_transf_8; 1.
SO SEQUENCE 290 AA; 33801 MW; 3A1A6FAF40C7AF7B CRC64;

Query Match 13.9%; Score 230.5; DB 2; Length 290;
Best Local Similarity 28.8%; Pred. No. 1,1e-12;
Matches 74; Conservative 46; Mismatches 100; Indels 37; Gaps 11;

OY 17 VAAKSEAAHPDTEIRFHVLDAGISEENRAAVANLRGG---NIRFI---DVPEDFA 69
   || : : : || : || : : : : : : : : : : : : : : : : : : : :
DB 5 VALYSIIENNHRINHFIHLISGI-EKECSAFYELGEPNTSISVYITDKFDINDTLV 63
   || : : : || : || : : : : : : : : : : : : : : : : : : : :

OY 70 -GEPINIHISTTYARKLGLEYIAD-CDKYVLYLDTDYLVRDGLKPLMDTDLGNNVGAC 127
   || : : : || : || : : : : : : : : : : : : : : : : : : : :
DB 64 LGIPL-----STCLRFLIPREVIDDKINKITLYLDCDVCNGPLDELVDYNLMAD-IACV 115
   || : : : || : || : : : : : : : : : : : : : : : : : : : :

OY 128 IDLFVEROEGYKOKIGMADGEYFNAGVLLINLKKRRHDIFFKMSCEWVEQYKDVMOYOD 187
   || : : : || : || : : : : : : : : : : : : : : : : : : : :
DB 116 IPDSPEMQERVKKLDYGLEFINFYFAGVAFINTSEKKNITQKALEMINSK-VYRYAD 174
   || : : : || : || : : : : : : : : : : : : : : : : : : : :

OY 188 ODILNGLFKGCVCYANSRFRNFPNTYAFMANGFASRHTDPLYLDRTNTAMP-VAVSHYCG 246
   || : : : || : || : : : : : : : : : : : : : : : : : : : :
DB 175 ODVLNLILNGRVHYLDKKYN-----NKTTLISVRCDEQKRLPNTIIMHYVT 220
   || : : : || : || : : : : : : : : : : : : : : : : : : : :

OY 247 SAKPWHRDCTWGAERF 263
   || : : : || : || : : : : : : : : : : : : : : : : : : : :
DB 221 QNKPWYK---IFRAQNF 234
   || : : : || : || : : : : : : : : : : : : : : : : : : : :

```

Search completed: December 2, 2002, 12:02:17
 Job time : 28.2849 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 34.9785 Seconds

(without alignments)
1283.801 Million cell updates/sec

Title: US-10-007-267-5

Perfect score: 1764

Sequence: 1 MQLPVSYLCAYNAKRYFAQ.....APQYRKILKMLRPMKYSY 337

Scoring table: BLOSUM62

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :
1: /SID2/gcgdata/geneseq/genescp-emb1/AA1980.DAT.*
2: /SID2/gcgdata/geneseq/genescp-emb1/AA1981.DAT.*
3: /SID2/gcgdata/geneseq/genescp-emb1/AA1982.DAT.*
4: /SID2/gcgdata/geneseq/genescp-emb1/AA1983.DAT.*
5: /SID2/gcgdata/geneseq/genescp-emb1/AA1984.DAT.*
6: /SID2/gcgdata/geneseq/genescp-emb1/AA1985.DAT.*
7: /SID2/gcgdata/geneseq/genescp-emb1/AA1986.DAT.*
8: /SID2/gcgdata/geneseq/genescp-emb1/AA1987.DAT.*
9: /SID2/gcgdata/geneseq/genescp-emb1/AA1988.DAT.*
10: /SID2/gcgdata/geneseq/genescp-emb1/AA1989.DAT.*
11: /SID2/gcgdata/geneseq/genescp-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/genescp-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/genescp-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/genescp-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/genescp-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/genescp-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/genescp-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/genescp-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/genescp-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/genescp-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/genescp-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/genescp-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/genescp-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1764	100.0	337	AA191314	N. gonorrhoeae gly
2	1761	99.8	337	AAW05579	Lipo-oligosacchari
3	1049	59.5	348	AA191311	N. gonorrhoeae gly
4	1046	59.3	348	AAW05576	Neisseria polysyl
5	1036	58.7	346	AAU72923	Neisseria meningit
6	260	14.7	50	AAW89332	Neisseria gonorrh
7	243	13.8	298	AAW86313	Putative glycosyl
8	223.5	12.7	332	AAW47426	EPSN. Lactococcus
9	219.5	12.4	274	AAW90151	C glutamicum prote
10	215	12.2	316	AAW54072	Enzyme EPS8 which

11	215	12.2	316	21	AAW43774
12	209	11.8	965	21	AAW96213
13	209	11.8	965	23	AAW47335
14	209	11.8	972	20	AAW06212
15	209	11.8	972	21	AAW43099
16	207.5	11.8	278	21	AAW68976
17	207.5	11.8	965	23	AAW47336
18	205.5	11.6	268	21	AAW54093
19	205.5	11.6	268	21	AAW43795
20	204	11.6	334	21	AAW68963
21	201.5	11.4	702	21	AAW96212
22	200.5	11.4	327	23	ABW48563
23	199	11.3	324	18	AAW14078
24	199	11.3	324	18	AAW21177
25	199	11.3	328	21	AAW81720
26	195	11.1	345	22	AAW90140
27	194	11.0	321	23	ABP26805
28	190	10.8	322	21	AAW68975
29	187.5	10.6	322	21	AAW68974
30	187	10.6	301	21	AAW97206
31	186.5	10.6	336	21	AAW54098
32	186.5	10.6	336	21	AAW43800
33	185.5	10.5	270	21	AAW97203
34	185.5	10.5	281	19	AAW61236
35	185.5	10.5	281	23	ABP54654
36	185.5	10.5	332	21	AAW68962
37	182.5	10.3	313	23	ABP27303
38	182	10.3	442	23	ABW54192
39	181	10.3	260	20	AAW88312
40	180.5	10.2	322	21	AAW54071
41	180.5	10.2	322	21	AAW43773
42	175.5	9.9	350	22	AAW96485
43	174	9.9	329	21	AAW68985
44	173.5	9.8	209	22	AAW47427
45	169	9.6	310	23	ABP25821

ALIGNMENTS

RESULT 1
AAW91314
ID AAW91314 standard; Protein: 337 AA.
XX
AC AAW91314;
DT 09-JUL-1996 (first entry)
XX
DE N. gonorrhoeae glycosyltransferase lgtD.
XX
KW Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;
KW vaccine.
XX
OS Neisseria gonorrhoeae strain F62.
XX
PN W09610086-A1.
XX
PD 04-APR-1996.
XX
PF 25-SEP-1995; 95WO-US12317.
XX
PR 26-SEP-1994; 94US-0312387.
XX
PA (UNIQ) UNIV ROCKEFELLER.
XX
PI Gotschlich EC;
XX
DR WPI: 1996-200924/20.
XX
DR N-PSDB; AAT14061.
XX
PT Nucleic acids encoding glycosyl transferase(s) - used in the
PT diagnosis of infection with Neisseria and for the biosynthesis of
PT oligosaccharide(s)

Amino acid sequenc
P. multocida chond
Pasteurella multoc
Pasteurella multoc
P. multocida hyalu
CpsII protein whic
Enzyme EPSH involv
Amino acid sequenc
Cps2K protein whic
P. multocida hyalu
Listeria monocylo
S. thermophilus exo
S. thermophilus exo
Streptococcus pneu
C glutamicum prote
Streptococcus poly
CpsII protein whic
Campylobacter jej
Enzyme EPSM involv
Amino acid sequenc
Campylobacter jej
S. pneumoniae pneu
S. pneumoniae SPl
Cps2J protein whic
Streptococcus poly
Lactococcus lactis
Sugar transferase
Enzyme EPS7 which
Amino acid sequenc
Putative glycosyl
Amino acid sequenc
EPSM. Lactococcus
Streptococcus poly

XX Claim 11: Fig 2e; 81pp; English.
PS
XX
CC 5 Glycosyltransferases (AAR91311-15) are products of the 1gt locus
CC (AAR14061) of *Neisseria gonorrhoeae* strain F62. Glycosyltransferase
CC 1gtD (AAR91314) can be obtd. by expression of the 1gt coding
CC sequence in recombinant host cells. A method for adding GalNAc
CC or GlcNAc beta1-3 to Gal comprises contacting a reaction mixture
CC congl. activated GalNAc or GlcNAc to an acceptor moiety comprising a
CC Gal residue in the presence of 1gtD. Oligosaccharides can be produced
CC that, when attached to non-toxic lipids, are useful for *Neisseria*
CC vaccine prepn. Blood group core oligosaccharides, and mimics of
CC lacto-N-neotetraose, gangliosides and saccharide portions of
CC globoglycolipids can also be produced using the enzymes.
XX
SQ Sequence 337 AA:

Query Match 100.0%; Score 1764; DB 17; Length 337;
Best Local Similarity 100.0%; Pred. No. 3.6e-178;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

OY 1 MOPVSVLICAVNAEKYFPAQSLAAVVGQTWNRNDILIVDGSSTDGTPAIAHRHQEDGR 60
DB 1 MOPVSVLICAVNAEKYFPAQSLAAVVGQTWNRNDILIVDGSSTDGTPAIAHRHQEDGR 60
OY 61 RIISNPRNLGFIASLNIIGDELAKSGGGEYIARTDADDIASPGMIKIVGEMEKDSIIA 120
DB 61 RIISNPRNLGFIASLNIIGDELAKSGGGEYIARTDADDIASPGMIKIVGEMEKDSIIA 120
OY 121 MGAMLEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIVAAPFGNPPIHNTMTIMRSVIDG 180
DB 121 MGAMLEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIVAAPFGNPPIHNTMTIMRSVIDG 180
OY 181 GLRFDPAYIHAEDYKFWYAGKIGRLAYTPPALVYKRFHODQTSKYNLQOORTAKIKE 240
DB 181 GLRFDPAYIHAEDYKFWYAGKIGRLAYTPPALVYKRFHODQTSKYNLQOORTAKIKE 240
OY 241 EIRAGWKAAGIAVAGDCNLNYGLKSTAYALYEKALSODIGCLRFLVEYFLSEKYSYL 300
DB 241 EIRAGWKAAGIAVAGDCNLNYGLKSTAYALYEKALSODIGCLRFLVEYFLSEKYSYL 300
OY 301 TDLDLFLTRVVRMKLFAAPQYRKILKKMLRPWKYRSY 337
DB 301 TDLDLFLTRVVRMKLFAAPQYRKILKKMLRPWKYRSY 337

RESULT 2
AAR06579 standard; Protein; 337 AA.
ID AAR06579;
AC AAR06579;
XX
DT 21-MAR-1997 (first entry)
XX
DE Lipo-oligosaccharide gene-encoded protein.
XX
XX Polyglycosyltransferase; N-acetylglucosaminyl transferase;
KM N-acetylglucosaminyl transferase; lipo-oligosaccharide.
XX
OS *Neisseria gonorrhoeae* ATCC 33084.
XX
XX WO9640971-A1.
XX
XX 19-DEC-1996.
XX
XX 03-JUN-1996; 96WO-US08323.
XX
XX 07-JUN-1995; 95US-0478140.
XX
XX (NEOS-) NEOS TECHNOLOGIES INC.
XX
XX Buczala SL, Johnson KF, Roth S;
PI
XX

DR WPI: 1997-052351/05.
DR N-PSDB: AAT9230.
XX
PT Transfer of at least 2 saccharide units using
PT poly:glycosyl:transferase - isolated from *N. gonorrhoeae*, catalyses
PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
PT single galactose moiety
XX
PS Disclosure: Fig 2E-F; 38pp; English.
XX
CC A lipo-oligosaccharide-encoding gene region (AAT9230) of *Neisseria*
CC *gonorrhoeae* ATCC 33084 includes coding sequences for 5 proteins
CC (AAR06576-80), one of which (AAR06576) is a polyglycosyltransferase
CC that catalyses the addition of GlcNAc and GalNAc disaccharides to
CC a galactose moiety. The function of the other 4 proteins is not
CC stated in the specification.
XX
SQ Sequence 337 AA:

Query Match 99.8%; Score 1761; DB 18; Length 337;
Best Local Similarity 99.7%; Pred. No. 7.5e-178;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

OY 1 MOPVSVLICAVNAEKYFPAQSLAAVVGQTWNRNDILIVDGSSTDGTPAIAHRHQEDGR 60
DB 1 MOPVSVLICAVNAEKYFPAQSLAAVVGQTWNRNDILIVDGSSTDGTPAIAHRHQEDGR 60
OY 61 RIISNPRNLGFIASLNIIGDELAKSGGGEYIARTDADDIASPGMIKIVGEMEKDSIIA 120
DB 61 RIISNPRNLGFIASLNIIGDELAKSGGGEYIARTDADDIASPGMIKIVGEMEKDSIIA 120
OY 121 MGAMLEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIVAAPFGNPPIHNTMTIMRSVIDG 180
DB 121 MGAMLEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIVAAPFGNPPIHNTMTIMRSVIDG 180
OY 181 GLRFDPAYIHAEDYKFWYAGKIGRLAYTPPALVYKRFHODQTSKYNLQOORTAKIKE 240
DB 181 GLRFDPAYIHAEDYKFWYAGKIGRLAYTPPALVYKRFHODQTSKYNLQOORTAKIKE 240
OY 241 EIRAGWKAAGIAVAGDCNLNYGLKSTAYALYEKALSODIGCLRFLVEYFLSEKYSYL 300
DB 241 EIRAGWKAAGIAVAGDCNLNYGLKSTAYALYEKALSODIGCLRFLVEYFLSEKYSYL 300
OY 301 TDLDLFLTRVVRMKLFAAPQYRKILKKMLRPWKYRSY 337
DB 301 TDLDLFLTRVVRMKLFAAPQYRKILKKMLRPWKYRSY 337

RESULT 3
AAR91311 standard; Protein; 348 AA.
ID AAR91311;
AC AAR91311;
XX
DT 09-JUL-1996 (first entry)
XX
XX N. gonorrhoeae glycosyltransferase 1gtA.
DE
XX Glycosyltransferase; lipo-oligosaccharide; 1gt gene; LOS locus;
KM vaccine.
XX
XX *Neisseria gonorrhoeae* strain F62.
XX
XX WO9610086-A1.
XX
XX 04-APR-1996.
XX
XX 25-SEP-1995; 95WO-US12317.
XX
XX 26-SEP-1994; 94US-0312387.
XX
XX (UYRQ) UNIV ROCKEFELLER.
XX
XX

PI Gotschlich EC;
 XX
 DR WPI: 1996-200924/20.
 XX N-PSDB: AAT14061.
 XX
 PT Nucleic acids encoding glycosyl transferase(s) - used in the
 PT diagnosis of infection with Neisseria and for the biosynthesis of
 PT oligo:saccharide(s)
 XX
 PS Claim 8: Fig 2b; 81pp; English.
 XX
 CC 5 Glycosyltransferases (AAR91311-15) are products of the 1gt locus
 CC (AAT14061) of Neisseria gonorrhoeae strain F62. Glycosyltransferase
 CC LgtA (AAR91311) can be obt. by expression of the 1gtA coding
 CC sequence in recombinant host cells. A method for adding GalNAc
 CC or GlcNAc beta1-3 to Gal comprises contacting a reaction mixture
 CC conng. activated GalNAc or GlcNAc to an acceptor molety comprising
 CC a Gal residue in the presence of LgtA. Oligosaccharides can be
 CC produced that, when attached to non-toxic lipids, are useful for
 CC Neisseria vaccine prepn. Blood group core oligosaccharides, and
 CC mimics of lacto-N-neotetraose, gangliosides and saccharide portions
 CC of globoglycolipids can also be produced using the enzymes.
 XX
 SQ Sequence 348 AA;
 Query Match 59.5%; Score 1049; DB 17; Length 348;
 Best Local Similarity 64.5%; Pred. No. 2.5e-102;
 Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;
 OY 1 MOPLVSVLICAYNAEKYFAQSLAAVVGOTWRNLDLIVDGGSTGTPATARRHFOEDGRI 60
 DB 1 MOPLVSVLICAYNAEKYFAQSLAAVVGOTWRNLDLIVDGGSTGTPATARRHFOEDGRI 60
 OY 61 RIISNPRNGFASINIGDELAKS--GGGEYIARTDADDIASPGIEKIVGEMEKDRSI 118
 DB 61 RIISNPRNGFASINIGDELAKS--GGGEYIARTDADDIASPGIEKIVGEMEKDRSI 120
 OY 119 IMGAWLEVLSEENKNSVAAIARNGAIWDKPTRHEDIYAAPFGNPINNTMIMRSYI 178
 DB 121 IMGAWLEVLSEENKNSVAAIARNGAIWDKPTRHEDIYAAPFGNPINNTMIMRSYI 180
 OY 179 DGLRFPDPAIHAEDYKFWYEAAGLGRALAYYPEALVKYRFHODQTSKKNLQOARTWAKI 238
 DB 181 DGLRFPDPAIHAEDYKFWYEAAGLGRALAYYPEALVKYRFHODQTSKKNLQOARTWAKI 240
 OY 239 KEETIRAGYKAAGIAGVADCLWYGLKSTAYALYKALSGOIGCLRFLYEFYLSLEY 298
 DB 241 KEETIRAGYKAAGIAGVADCLWYGLKSTAYALYKALSGOIGCLRFLYEFYLSLEY 300
 OY 299 SLTDLDFLTDRVMRKLEFAAPQYRKIKKMLR 330
 DB 301 PSGAMLDFAADGMRRLFTLROYFGILYRLIK 332
 RESULT 4
 AAM06576
 ID AAM06576 standard; Protein; 348 AA.
 XX
 AC AAM06576;
 XX
 DT 21-MAR-1997 (first entry)
 XX
 DE Neisseria polyglucosyltransferase.
 XX
 KM Polyglucosyltransferase: N-acetylglucosaminyl transferase;
 KM N-acetylglucosaminyl transferase: lipo-oligosaccharide.
 OS Neisseria gonorrhoeae ATCC 33084.
 XX
 PN W09640971-A1.
 XX
 PD 19-DEC-1996.
 XX

PF 03-JUN-1996; 96WO-US08323.
 XX
 PR 07-JUN-1995; 95US-0478140.
 XX
 XX (NEOS-) NEOSE TECHNOLOGIES INC.
 PA Buczala SL, Johnson KF, Roth S;
 PI WPI: 1997-052351/05.
 XX N-PSDB: AAT49230.
 DR
 XX
 PT Transfer of at least 2 saccharide units using
 PT poly:glycosyltransferase - isolated from N. gonorrhoeae, catalyses
 PT the addition of both GlcNAc and GalNAc di:saccharide(s) units to a
 PT single galactose molety
 XX
 PS Disclosure: Fig 2A-C; 38pp; English.
 XX
 CC A novel polyglucosyltransferase (PGTase) (AAM06576) from Neisseria
 CC gonorrhoeae ATCC 33084 catalyses the stereospecific conjugation of
 CC 2 specific activated saccharide units (e.g. UDP-GlcNAc, UDP-GalNAc,
 CC UDP-Gal) to specific acceptors having a galactose molety at a
 CC non-reducing end. It is the first PGTase reported to be capable of
 CC transfer of more than one different saccharide molety. The PGTase
 CC is encoded by nucleotides 445-1488 of a lipo-oligosaccharide gene
 CC (AAT49230). It can be produced in transformed host cells and used in
 CC oligosaccharide prodn.
 XX
 SQ Sequence 348 AA;
 Query Match 59.3%; Score 1046; DB 18; Length 348;
 Best Local Similarity 64.2%; Pred. No. 5.3e-102;
 Matches 213; Conservative 34; Mismatches 83; Indels 2; Gaps 1;
 OY 1 MOPLVSVLICAYNAEKYFAQSLAAVVGOTWRNLDLIVDGGSTGTPATARRHFOEDGRI 60
 DB 1 MOPLVSVLICAYNAEKYFAQSLAAVVGOTWRNLDLIVDGGSTGTPATARRHFOEDGRI 60
 OY 61 RIISNPRNGFASINIGDELAKS--GGGEYIARTDADDIASPGIEKIVGEMEKDRSI 118
 DB 61 RIISNPRNGFASINIGDELAKS--GGGEYIARTDADDIASPGIEKIVGEMEKDRSI 120
 OY 119 IMGAWLEVLSEENKNSVAAIARNGAIWDKPTRHEDIYAAPFGNPINNTMIMRSYI 178
 DB 121 IMGAWLEVLSEENKNSVAAIARNGAIWDKPTRHEDIYAAPFGNPINNTMIMRSYI 180
 OY 179 DGLRFPDPAIHAEDYKFWYEAAGLGRALAYYPEALVKYRFHODQTSKKNLQOARTWAKI 238
 DB 181 DGLRFPDPAIHAEDYKFWYEAAGLGRALAYYPEALVKYRFHODQTSKKNLQOARTWAKI 240
 OY 239 KEETIRAGYKAAGIAGVADCLWYGLKSTAYALYKALSGOIGCLRFLYEFYLSLEY 298
 DB 241 KEETIRAGYKAAGIAGVADCLWYGLKSTAYALYKALSGOIGCLRFLYEFYLSLEY 300
 OY 299 SLTDLDFLTDRVMRKLEFAAPQYRKIKKMLR 330
 DB 301 PSGAMLDFAADGMRRLFTLROYFGILYRLIK 332
 RESULT 5
 AAU72923
 ID AAU72923 standard; Protein; 346 AA.
 XX
 AC AAU72923;
 XX
 DT 12-MAR-2002 (first entry)
 XX
 DE Neisseria meningitidis virulence protein #13.
 XX
 KM Meningitidis: virulence; gene: antibacterial; vaccine; veterinary;
 KM infection; Gram-negative bacteria; antimicrobial.
 OS Neisseria meningitidis.
 OS

```
XX PN WO200185772-A2.
XX PD 15-NOV-2001.
XX PF 08-MAY-2001; 2001WO-GB02003.
XX PR 08-MAY-2000; 2000GB-0011108.
XX PA (MICR-) MICROSCIENCE LTD.
XX PI Tang C;
XX DR WPI; 2002-066593/09.
XX DR N-PSDB; AAS97208.
XX PT New peptide encoded by operon including virulence genes of Neisseria
XX PT meningitidis, useful as vaccine component for treating or preventing
XX PT meningitis and for identifying antimicrobial drug
XX PS Claim 4; Page 60-61; 423pp; English.
XX CC The invention relates to a peptide (I) encoded by an operon (II) of
XX CC Neisseria meningitidis including virulence genes, or a related molecule
XX CC having a 40% sequence similarity at the peptide or nucleotide level in a
XX CC Gram-negative bacterium, or its functional fragment, for therapeutic or
XX CC diagnostic use. (I) and (II) are useful in the manufacture of a
XX CC medicament for treating or preventing a condition (e.g., meningitis)
XX CC associated with infection by Neisseria or Gram-negative bacteria. The
XX CC product is useful for veterinary treatment and in a screening assay for
XX CC the identification of an antimicrobial drug. The vaccines have
XX CC CC propylactic applications. AAU72911-AAU73014 represent N. meningitidis
XX CC virulence proteins of the invention.
SQ Sequence 346 AA:
Query Match 58.7%; Score 1036; DB 23; Length 346;
Best Local Similarity 63.0%; Pred. No. 6e-101;
Matches 208; Conservative 36; Mismatches 86; Indels 0; Gaps 0;
QY 1 MOPLVSLICAVYAEKVFQSLAAVVGQTRNLDLIVDGSNDGTPALRHOEDGRI 60
DB 1 LQPLVSLICAVYAEKVFQSLAAVVGQTRNLDLIVDGSNDGTPALRHOEDGRI 60
QY 61 RIISNPNLGFIAISLNLGDELAKSGGGEYIARTDADIASPGMIKIVGEMEKRSIIA 120
DB 61 KILAAQNSGLIPSLNIGDELAKSGGGEYIARTDADIASPGMIKIVGEMEKRSIIA 120
QY 121 MGAWLEVLSSENNKSVLAIAIRNGATWDKPTRHEDIYAVFPFGNPINNTMIRRSYIDG 180
DB 121 MGAWLEVLSSENNKSVLAIAIRNGATWDKPTRHEDIYAVFPFGNPINNTMIRRSYIDG 180
QY 181 GLFEDPAIYAEEDYKFEYKGLRLAYVPEALVYKRFHDDTSKYNLQORRTAWKIKE 240
DB 181 GLRYNTERMAEDYQFYWDVSKLGRALAYPEALVYKRLAHNOVSKSYIRQHEIAQGIQK 240
QY 241 EIRAGYKAGIAGVADCLNYGLKSTAYALYERKALSGDIGCLRLFYFLSLEKYSYL 300
DB 241 TARNDFLQSGWGFTRFDSLETRQIKAVAYELLEKHLPEEDFERARRLVCGCFRTDTPPA 300
QY 301 TDLIDFLTRVYMKRLFAAPYRKILKKMLR 330
DB 301 GAWLIDFAADGKMRLFTMRQYFGLHRLIK 330
RESULT 6
ID AAW89332 standard; peptide; 50 AA.
XX AAW89332;
AC AAW89332;
XX 26-FEB-1999 (first entry)
XX
```

```
DE Neisseria gonorrhoeae IgtD C-terminal peptide.
XX Neisseria meningitidis; IgtC; IgtB; beta-1,4-galactosyltransferase;
XX glycosyltransferase; proteolytic enzyme.
XX Neisseria gonorrhoeae.
XX WO9854331-A2.
XX PD 03-DEC-1998.
XX PF 26-MAY-1998; 98WO-IB00975.
XX PR 27-MAY-1997; 97US-0047751.
XX PA (CANA ) NAT RES COUNCIL CANADA.
XX PI Wakarchuk WM, Young NM;
XX DR WPI; 1999-035177/03.
XX PT Expressing high levels of glycosyltransferases - comprises use of
XX PT either host cells deficient in proteolytic enzymes or modified
XX PT glycosyltransferase genes deleted in a proteolytic recognition site
XX PS Example 1; Fig 8; 61pp; English.
XX CC A method has been developed of expressing a glycosyltransferase in a
XX CC host cell. The method comprises introducing into the host cell a nucleic
XX CC acid encoding the glycosyltransferase and incubating the host cell under
XX CC conditions appropriate for expression of the glycosyltransferase, where
XX CC the host cell substantially lacks a protease that cleaves polypeptides
XX CC between two consecutive positively charged amino acid residues. The
XX CC glycosyltransferase can be used in in vitro production of
XX CC oligosaccharide structures which are potential therapeutic agents for
XX CC use in the manipulation of cell-cell recognition events, particularly
XX CC adhesion of bacteria and viruses to mammalian cells and leukocyte-
XX CC endothelial cell interaction through selectins in inflammation. The
XX CC method provides more readily recoverable active glycosyltransferases
XX CC than prior art methods involving mammalian glycosyltransferases. The
XX CC present sequence represents a C-terminal peptide from Neisseria
XX CC gonorrhoeae IgtD from the present invention.
SQ Sequence 50 AA:
Query Match 14.7%; Score 260; DB 20; Length 50;
Best Local Similarity 100.0%; Pred. No. 6.2e-20;
Matches 50; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
QY 285 RLFLYEFSLSEKYSLTDLDFLTRVYMKRLFAAPYRKILKKMLRPMKY 334
DB 1 RLFLYEFSLSEKYSLTDLDFLTRVYMKRLFAAPYRKILKKMLRPMKY 50
RESULT 7
ID AAB96313 standard; Protein; 298 AA.
XX AAB96313;
AC AAB96313;
XX 29-OCT-2001 (first entry)
XX DE Putative glycosyltransferase, involved in cell wall biogenesis #1.
XX KW Hyperthermophilic archaeon; hyperthermophilic protein.
XX OS Pyrococcus abyssi.
XX PN FR2792651-A1.
XX PD 27-OCT-2000.
XX PF 21-APR-1999; 99FR-0005034.
```

[illegible]

```

XX XX WO200157234-A2.
XX PD 09-AUG-2001.
XX PF 02-FEB-2001; 2001WO-US03404.
XX PR 02-FEB-2000; 2000US-0179888.
XX PR 16-OCT-2000; 2000US-0241098.
XX PA (UYOR-) UNIV OREGON STATE.
XX PA (USDA) US DEPT OF AGRICULTURE.
XX PL Trempey JE, Knoshaug EP, Sandine WE, Ahlgren JA, Dierksen KP;
XX DR WPI; 2001-488889/53.
XX DR N-PSDB; AAH43198.
XX PT New bacterium useful in pharmaceutical formulations, food products and
XX PT beauty cosmetics; comprises characteristics of Lactococcus lactis
XX PT cremoris Ropy 352.
XX PS Claim 28; Page 67-68; 73pp; English.
XX CC The sequences given in AAH47426-27 are encoded by a fragment of the EPS
XX CC plasmid derived from L. lactis subspecies cremoris Ropy352. The EPS
XX CC plasmid is about 32 kb in size and encodes at least 13 active genes.
XX CC The enzymes encoded by these genes allow the bacteria to produce an
XX CC exopolysaccharide, designated EPS352. When EPS352 is expressed in or
XX CC added to milk, it imparts highly desirable sensory characteristics to
XX CC the milk, including making the milk very thick, with a very smooth
XX CC mouth-feel, and slightly sweet with an obvious chewable-bite. Open
XX CC reading frames (ORF's) M and N show homology to glycosyltransferase
XX CC involved in EPS352 biosynthesis. L. lactis cremoris Ropy352 is
XX CC deposited with the USDA-ARS-NCMUR-NRRL as deposit accession number
XX CC NRRL B-30729. EPS352 is useful for thickening a liquid selected from
XX CC milk, a milk-based liquid, a whey-based liquid, a soy-based liquid,
XX CC and a fruit-juice. It is also useful as an additive in pharmaceutical
XX CC products, beauty care products and coating agents.
XX SQ Sequence 332 AA:
XX
XX Query Match 12.7%; Score 223.5; DB 22; Length 332;
XX Best Local Similarity 24.8%; Pred. No. 8,4e-15;
XX Matches 61; Conservative 55; Mismatches 81; Indels 49; Gaps 8;
XX
XX QY 4 LYSVLICANNAKRYFAQSLAAVVGQTRWRDLILYDGDSTGTPAARHFOGODGIRII 63
XX LI::II::II::II::II::II::II::II::II::II::II::II::II::II::
XX 5 LII::IIVPNVNSKRYRAALHSLNDLTQYIEVILLINDGSTDSQELISSFOKKDKRIKY 64
XX
XX QY 64 SNPRNLGFASINIGDELAKSGGGEYIARTADDIASPGWIEKIVGEMEKRSITAMGA 123
XX I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::
XX DB 65 -NTRKULGVSHANNYIDR-----ASSGYIEFLDPDDTIDYDKSYCLEMGLINKRNADYVMN 119
XX
XX QY 124 W-----LE---VLSEENN-KSVLAIAIARNGAIDMKPTRHEDIIVAVFPG 163
XX I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::
XX DB 120 YXICKGNITYPVWNNDLLECEGLLSRDKTMRISILSDPTGKGFYW----- 163
XX
XX QY 164 NPIHNNTMIMRSVYIDGILRDPDPAITHADYKFWYFAGLGR-LAYYPALKYRPHOD 222
XX I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::I::
XX DB 164 -----TRIFRRKNVIN-NVKFMESINYLEDMFNISIVNARIATYNRKHYLORES 216
XX
XX QY 223 TSSKYN 228
XX I::I::
XX DB 217 ASKFFS 222
XX
XX RESULT 9
XX AAG90151
XX ID AAG90151 standard; Protein; 274 AA.
XX AC AAG90151;
XX

```

DT 26-SEP-2001 (first entry)
XX C glutamicum protein fragment SEQ ID NO: 3905.
XX
XX Coryneform bacterium: amino acid synthesis; vitamin; saccharide;
KM organic acid synthesis.
XX Corynebacterium glutamicum.
OS
XX
XX EPI108790-A2.
PN
XX
XX 20-JUN-2001.
PD
XX
XX 18-DEC-2000; 2000EP-0127688.
PF
XX 16-DEC-1999; 99JP-0377484.
PR 07-APR-2000; 2000JP-0159162.
PR 03-AUG-2000; 2000JP-0280988.
XX
XX (KYOW) KYOWA HAKKO KOGYO KK.
PA
XX
XX Nakagawa S, Mizoguchi H, Ando S, Hayashi M, Ochiai K, Yokoi H;
PI Tateishi N, Senoh A, Ikeda M, Ozaki A;
XX
XX WPI; 2001-376931/40.
DR N-PSDB; AAH5370.
XX
XX Novel polynucleotides derived from Coryneform bacteria, for identifying
PT mutation point of a gene, measuring expression of a gene, analysing
PT expression profile or pattern of a gene and identifying homologous gene
PT
XX
XX
XX Claim 17; SEQ ID NO: 3905; 246pp + Sequence Listing; English.
XX

CC The present invention provides a number of nucleotide and protein
CC sequences from the Coryneform bacterium Corynebacterium glutamicum. These
CC are useful for identifying the mutation point of a gene derived from a
CC mutant of coryneform bacterium, measuring expression amount and
CC analysing the expression profile or expression pattern of a gene derived
CC from Coryneform bacterium, and identifying a homologue of a gene derived
CC from coryneform bacterium. Coryneform bacteria are useful for producing
CC amino acids, nucleic acids, vitamins, saccharides and organic acids,
CC particularly L-lysine. The present sequence is a protein described
CC in the exemplification of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in electronic format directly from the
CC European Patent Office.
XX
XX

SO Sequence 274 AA;

Query Match 12.4%; Score 219.5; DB 22; Length 274;
Best Local Similarity 27.1%; Pred. No. 1.7e-14;
Matches 78; Conservative 57; Mismatches 108; Indels 45; Gaps 13;

QY 3 PLYSVLICAVNAEKYFQSLAAVVGOTWRNLDLIYDGGSTGTPA-IARHQEDGRT 60
DB 6 PLISVVIPTLAYEYCSQSIKVCSEONYENWQIVLV---LDCAPIKDVPQWKEHE-RI 60
QY 61 RIISNRFNFIASININGIDELAKSSGGEYIARTDADDIASPMIRIKYEMEKKSIIA 120
DB 61 KIYEQXIRCTPTSLNNGI---KASDGLIARLSDSDLAAPSRLSQEFELNNHPYITC 116
QY 121 MGAMLEVLSEENKSVLAIAIRNGAIV---DKPTRHEDIVAVFPNGPIHNTMIMRRS 176
DB 117 VAKTKHINE-----HGKIFGOSADLPT--SQDIRQLLVKNPITHSVWRKQ 163
QY 177 VID--GGLRPDPAYIAEDYKFWYEGAKGLRLAYYPDALVKYFPHODQSSKYNLQORT 234
DB 164 VYEOIGYSLE--MTRSODYELFLRLSIGALIGYLDSELSYSYRHHGQHSRKTSPEPKYT 221
QY 235 AMKIKKEIR-AGYWAAGTA-----VGADCLNY-GILKS---TAY 269
DB 222 WILKRMELASPLKRSVPRQIFLNFITWYGAOVTRYLGLKRAKGFMTGY 269

RESULT 10

AA54072
ID AA54072 standard; Protein; 316 AA.

AA54072;

DT 27-MAR-2000 (first entry)

DE Enzyme EPS8 which is involved in exopolysaccharide biosynthesis.

XX
XX Exopolysaccharide; EPS; ESP enzyme; EPS1; EPS2; EPS3; EPS4; EPS5;
KM EPS6; EPS7; EPS8; EPS9; EPS10; Streptococcus thermophilus strain SF139;
KM activated D-galactose pyranose; saccharide; beta-glycosyltransferase;
KM transporter; food; fermented milk product; yoghurt; cheese;
KM flavour stability; organoleptic property.
XX
XX Streptococcus thermophilus.
OS
XX
XX MO9962316-A2.
PN
XX
XX 09-DEC-1999.
PD
XX
XX 22-APR-1999; 99WO-EP02841.
PF
XX
XX 22-APR-1998; 98EP-0201310.
PR 22-APR-1998; 98EP-0201311.
PR 22-APR-1998; 98EP-0201312.
XX
XX (NEST) SOC PROD NESTLE SA.
PA
XX
XX Stinglele F, Germond JE, Lamothe G;
PI
XX

DR WPI: 2000-097267/08.
DR N-PSDB; AA245258, AA54072, AA54073, AA54074.

PT New recombinant enzymes for synthesis of exopolysaccharides,
PT particularly in lactic acid bacteria, for improving properties of
PT fermented milk products
XX
XX
XX Claim 3; Page 105-106; 162pp; French.

XX
XX AA54065-74 represent enzymes involved in the biosynthesis of
XX exopolysaccharides (EPS). These enzymes are designated EPS1-EPS10. and
XX are encoded by open reading frames eps1-eps10. The enzymes are isolated
XX from Streptococcus thermophilus strain sf139. The proteins are used
XX in a method for the synthesis of EPS, which includes at least one step
XX of forming a bond (alpha or beta-isomer) between C-1 (carrying the
XX reducing aldehyde function, of an activated D-galactose pyranose), and
XX a phosphate on a lipophilic or proteinaeous primer. Stepwise synthesis
XX of EPS occurs with, in each step, addition of a new sugar unit, through
XX its hemi-acetyl function, to an alcoholic hydroxyl of a second sugar
XX unit, present at the end of a chain of sugar residues bonded to the
XX primer. EPS1 to EPS4 are used to elongate the EPS chain, and to regulate
XX EPS production, EPS5 creates new bonds between saccharides, EPS6 is used
XX in the biosynthesis of EPS, EPS7 and EPS8 are beta-glycosyltransferases,
XX EPS9 is a transporter of repetitive units, and EPS10 catalyses the
XX conversion of a pyranose form of a beta-D-galactose to the furanose
XX form. The EPS enzyme are used to improve properties of foods,
XX particularly fermented milk products such as yoghurt and cheese,
XX e.g. their organoleptic properties and flavour stability.

SO Sequence 316 AA;

Query Match 12.2%; Score 215; DB 21; Length 316;
Best Local Similarity 23.1%; Pred. No. 6.2e-14;
Matches 81; Conservative 59; Mismatches 105; Indels 106; Gaps 14;

QY 1 MOPLYSVLICAVNAEKYFQSLAAVVGOTWRNLDLIYDGGSTGTPA-IARHQEDGRT 60
DB 1 MNPISIVPTVYVKEVYIRCTIESIIAQYIRNLEVIIVDGGSTDSLAIVISDLICSHNI 60


```

OY 61 RIISNPNLGFASLNLIGLDELAKSGGGEYIARTDADDIASPGMIKEIYGEKEDRSITA 120
DB 61 KYI-MOKNOGLSVARNTGID----AATGKXIAFVDADDXIKPDPVSLQIADKTGADIV 115
OY 121 MGAMLEVLSEENKSVLAIAIRNGAI-----WDKPTRHEDIIVAFPGNPPIHNN----- 169
DB 116 RGSFRRF-----NGNIPKGMVDPDENVPNTVGTIV-----LDQFLSSNISFV 156
OY 170 --TMIRRSVYIDGG-LRPDPAYIHAEDYKFWYEAGKLGRL-AVYEPALVKYRFHOD----- 221
DB 157 VWSISYIRLDPFINSNHRFPGLIF-EDADFTIRAYMLAKLVATSPPNAYRINRPGSITL 215
OY 222 QTSKXKYNLQORRTAWKIKEIRAGYKWAAGIANGADCLNGLKSTAYALYEKALSODI 281
DB 216 TTKTKYNAOK-----MSLSEKIIISQ--- 236
OY 282 GCLRFLFYELFSLKYSITLDLDFLTDRVMKRLFAAPQYRKILKMLRPW 332
DB 237 -----FISMLKHEKSDVLCSL--ILKSIYAC-----MDMW 264

RESULT 11
AAV43774
ID AAV43774 standard; Protein; 316 AA.
AC AAV43774;
XX
XX 11-FEB-2000 (first entry)
DE Amino acid sequence of eps8 of Streptococcus thermophilus Sfi39.
XX
XX eps operon; Streptococcus thermophilus Sfi39; enzyme; eps1; eps2;
XX eps3; eps4; eps5; eps6; eps7; eps8; eps9; eps10;
XX exopolysaccharide biosynthesis; EPS; intersugar bond; antitumour;
XX probiotic; foodstuff; organoleptic quality; flavour;
XX lactic acid bacteria; acidified milk product; yoghurt; cheese.
XX
XX Streptococcus thermophilus.
OS
XX
XX W09954475-A2.
PN
XX
XX 28-OCT-1999.
PD
XX
XX 22-APR-1999; 99WO-EP03011.
PF
XX
XX 22-APR-1998; 98EP-0201310.
PR 22-APR-1998; 98EP-0201311.
PR 22-APR-1998; 98EP-0201312.
XX
XX (NEST ) SOC PROD NESTLE SA.
PA
XX
XX Stinglele F, Germond JE, Lamothe G;
PI
XX
XX WPI; 2000-013255/01.
DR N-PSDB; AA230355, AAV43774, AAV43775, AAV43776;
XX
XX New recombinant enzymes for biosynthesis of exopolysaccharides having
XX e.g. antitumor or probiotic properties or useful in fermented milk
XX products -
XX
XX Claim 3; Page 106-107; 163pp; French.
XX
XX AAV43767-76 represent the proteins encoded by the eps operon of
XX Streptococcus thermophilus Sfi39. The operon contains 10 open reading
XX frames, and encodes enzymes (eps1, eps2, eps3, eps4, eps5, eps6, eps7,
XX eps8, eps9 and eps10) that are involved in the biosynthesis of
XX exopolysaccharides (EPS). The enzymes catalyse the formation of
XX specific intersugar bonds. The enzymes catalyse a process that includes
XX at least one step of forming a bond (in alpha or beta anomeric form)
XX between C1, carrying the reducing aldehyde group of an activated D-galp
XX (galactose in pyranose form), and a phosphate group on a lipophilic or
XX proteinaceous primer. The enzymes are used to produce EPS that have
XX antitumor or probiotic properties or are used in foodstuffs to improve
XX

```

```

CC organoleptic qualities and flavour. When expressed by lactic acid
CC bacteria, EPS impart a free-flowing character and/or a smooth, creamy
CC texture to acidified milk products (yoghurt or cheese).
XX
XX
SQ Sequence 316 AA:
Query Match 12.2%; Score 215; DB 21; Length 316;
Best Local Similarity 23.1%; Pred. No. 6, 2e-14;
Matches 81; Conservative 59; Mismatches 105; Indels 106; Gaps 14;

OY 1 MOPVSVLCAINAEKFFQSLAAVVGQTRNLDILIVDGGSDIGTPALARHQEDGRI 60
DB 1 MNPILSIIVPTYNVEKYIRTCIESIIAQYRNIEVIIVDGGSDIOLASIDLCISHNI 60
OY 61 RIISNPNLGFASLNLIGLDELAKSGGGEYIARTDADDIASPGMIKEIYGEKEDRSITA 120
DB 61 KYI-MOKNOGLSVARNTGID----AATGKXIAFVDADDXIKPDPVSLQIADKTGADIV 115
OY 121 MGAMLEVLSEENKSVLAIAIRNGAI-----WDKPTRHEDIIVAFPGNPPIHNN----- 169
DB 116 RGSFRRF-----NGNIPKGMVDPDENVPNTVGTIV-----LDQFLSSNISFV 156
OY 170 --TMIRRSVYIDGG-LRPDPAYIHAEDYKFWYEAGKLGRL-AVYEPALVKYRFHOD----- 221
DB 157 VWSISYIRLDPFINSNHRFPGLIF-EDADFTIRAYMLAKLVATSPPNAYRINRPGSITL 215
OY 222 QTSKXKYNLQORRTAWKIKEIRAGYKWAAGIANGADCLNGLKSTAYALYEKALSODI 281
DB 216 TTKTKYNAOK-----MSLSEKIIISQ--- 236
OY 282 GCLRFLFYELFSLKYSITLDLDFLTDRVMKRLFAAPQYRKILKMLRPW 332
DB 237 -----FISMLKHEKSDVLCSL--ILKSIYAC-----MDMW 264

RESULT 12
AAV96213
ID AAV96213 standard; Protein; 965 AA.
XX
XX
XX AAV96213;
AC
XX
XX 17-AUG-2000 (first entry)
DE
XX
XX P. multocida chondroitin synthase.
DE
XX
XX Chondroitin synthase; CS; enzyme; hyaluronic acid; ulcer;
XX tissue abrasion; viscoelastic replacement; bioadhesive.
XX
XX Pasteurella multocida.
OS
XX
XX W0200027437-A2.
PN
XX
XX 18-MAY-2000.
PD
XX
XX 10-NOV-1999; 99WO-US26501.
PF
XX
XX 11-NOV-1998; 98US-0107929.
PR 01-APR-1999; 99US-0283402.
XX
XX (OKLA ) UNIV OKLAHOMA STATE.
PA
XX
XX Deangelis PL;
PI
XX
XX WPI; 2000-376319/32.
DR N-PSDB; AAA27449.
XX
XX Novel method for the enzymatic transfer of sugar molecules to an
XX acceptor, useful for synthesis of e.g. polysaccharide bioadhesives or
XX drug delivery systems, including hybrid molecules -
XX
XX Claim 15; Page 85; 86pp; English.
XX
XX The present sequence is the Pasteurella multocida chondroitin synthase
XX

```

CC PMCS. PMCS catalyzes glycosaminoglycan polymerisation to produce
 CC chondroitin: a linear polysaccharide which has viscoelastic properties
 CC which makes it useful for a number of applications. Chondroitin can be
 CC used with hyaluronic acid (HA) to coat medical devices e.g. catheters and
 CC sensors to reduce tissue abrasion. In addition, they can be used as
 CC bioadhesives for haemostatic sealing and healing of wounds and surgical
 CC incisions; and as biomaterials that provide sustained delivery of
 CC encapsulated drugs, to wounds, ulcers, injuries or surgical sites.

XX Sequence 965 AA:

Query Match 11.8%; Score 209; DB 21; Length 965;
 Best Local Similarity 23.7%; Pred. No. 1.5e-12;
 Matches 75; Conservative 53; Mismatches 129; Indels 60; Gaps 9;

QY 3 PLVSVLICAVNAEKYFAQSIAAVVGTWRNLDLIYDDSGTGPALAHFQDGRIRI 62
 DB 433 PLVSVIIPAINCANYIORCVDSALNOTVVDLEVCICNDGSTDWTLYNKLYGNRPVR 492
 QY 63 ISNPRNLGFASINIGDELAKSGGGEYIARTDADIASPGWIEKIVGEMKDRSIAMG 122
 DB 493 MSKPN--GGIASASNAVSPAK--GYIGQLSDDYLEPDVAELCKEFLKDKTLAC-- 545
 QY 123 AMLEVLSEENKSVLAIAIRNGAIPDKPTREHEDIVAVFPFGNPIHNNMTIMRS--VIDG 180
 DB 546 ----VTTNNVNPDSGLANGYNWPEFSREKLTAMI-----AHFRMTIIRAMHLTDG 596
 QY 181 GLRPDPAVIAHEDYKFWYEGKIGRLAYYPEALVKYRFHODOTSSKYNLQORRTAMKIKE 240
 DB 597 ----FNEINENAVDYDMFLKLESEVKRKLKICYNRYLHODNTSIRK----- 639
 QY 241 EIRAGYWKAGIAVGADCLMNYGLKSTAYALYERKALSGODIGCLRFLYEFYLSL--K 297
 DB 640 -----KLGIOKKNHFVYVNSLNROGI---NYYNKDFDLDDESRR 677
 QY 298 YSLTDLDFLTDRVMRK 314
 DB 678 YIFNKTAEYQEMDMRK 694

RESULT 13

AA047335
 ID AA047335 standard; Protein: 965 AA.

XX AA047335;

DT 22-FEB-2002 (first entry)

DE Pasteurella multocida chondroitin synthase #1.

XX Chondroitin synthase; dermatan sulphate; chondroitin sulphate polymer;
 KW eye application; joint application; moisturiser; drug delivery;
 KW wound dressing; biocompatible film.

XX Pasteurella multocida.

XX WO200180810-A2.

PD 01-NOV-2001.

XX 25-APR-2001; 2001WO-US13395.

XX 25-APR-2000; 2000US-199538P.

XX (DANG/) DE ANGELIS P L.

XX De Angelis PL;

XX MPI; 2002-049237/06.

XX N-PSDB; ABA05097.

XX New chondroitin synthase gene obtained from Pasteurella multocida,
 PT useful as hyaluronan polysaccharide substitute in medial or cosmetic

PT applications, e.g. for eye or joint applications, for moisturizer or
 PT wound dressings -
 XX
 XX
 PS Claim 3: Page 119; 125pp; English.

CC The present invention relates to the coding sequence of the Pasteurella
 CC multocida chondroitin synthase. A chondroitin polysaccharide may be used
 CC as a hyaluronan polysaccharide substitute in medial or cosmetic
 CC applications, for example in eye or joint applications, for moisturiser
 CC or wound dressings. The enzyme may be used in covalently coupling
 CC specific drugs, proteins or toxins to the structurally modified
 CC chondroitin for general or targeted drug delivery or radiological
 CC procedures, covalently cross linking the hyaluronic acid itself or to
 CC other supports to achieve a gel or other three dimensional biomaterial
 CC with stronger physical properties, and covalently linking hyaluronic acid
 CC to a surface to create a biocompatible film or monolayer. The present
 CC sequence is one version of the protein of the invention.

XX Sequence 965 AA:

Query Match 11.8%; Score 209; DB 23; Length 965;
 Best Local Similarity 23.7%; Pred. No. 1.5e-12;
 Matches 75; Conservative 53; Mismatches 129; Indels 60; Gaps 9;

QY 3 PLVSVLICAVNAEKYFAQSIAAVVGTWRNLDLIYDDSGTGPALAHFQDGRIRI 62
 DB 433 PLVSVIIPAINCANYIORCVDSALNOTVVDLEVCICNDGSTDWTLYNKLYGNRPVR 492
 QY 63 ISNPRNLGFASINIGDELAKSGGGEYIARTDADIASPGWIEKIVGEMKDRSIAMG 122
 DB 493 MSKPN--GGIASASNAVSPAK--GYIGQLSDDYLEPDVAELCKEFLKDKTLAC-- 545
 QY 123 AMLEVLSEENKSVLAIAIRNGAIPDKPTREHEDIVAVFPFGNPIHNNMTIMRS--VIDG 180
 DB 546 ----VTTNNVNPDSGLANGYNWPEFSREKLTAMI-----AHFRMTIIRAMHLTDG 596
 QY 181 GLRPDPAVIAHEDYKFWYEGKIGRLAYYPEALVKYRFHODOTSSKYNLQORRTAMKIKE 240
 DB 597 ----FNEINENAVDYDMFLKLESEVKRKLKICYNRYLHODNTSIRK----- 639
 QY 241 EIRAGYWKAGIAVGADCLMNYGLKSTAYALYERKALSGODIGCLRFLYEFYLSL--K 297
 DB 640 -----KLGIOKKNHFVYVNSLNROGI---NYYNKDFDLDDESRR 677
 QY 298 YSLTDLDFLTDRVMRK 314
 DB 678 YIFNKTAEYQEMDMRK 694

RESULT 14

AA06212
 ID AA06212 standard; Protein: 972 AA.

XX AA06212;

DT 16-AUG-1999 (first entry)

DE Pasteurella multocida hyaluronate synthase pmHAS.

XX Hyaluronate synthase; pmHAS; hyaluronic acid; hyaluronan.

XX Pasteurella multocida.

XX WO923227-A2.

XX 14-MAY-1999.

XX 30-OCT-1998; 98WO-US23153.

XX 26-OCT-1998; 98US-0178851.

XX 31-OCT-1997; 97US-0064435.

XX (OKLA) UNIV OKLAHOMA STATE.

```
XX Deangelis P, Kumari K, Weigel PH;
XX
XX WPI: 1999-337486/28.
DR N-PSDB: AAX58857.
XX
XX Nucleic acid encoding hyaluronate synthase for production of
XX hyaluronic acid with controlled molecular weight and targeting
XX specificity
XX
XX Disclousure: Page 123-125; 125pp; English.
XX
XX This present sequence represents the hyaluronate synthase (pmHAS)
XX of Pasteurella multocida Carter Type A. The pmHAS enzyme has
XX different kinetic optima with respect to pH and metal ion
XX dependence, and different Km values compared with the HAS enzymes
XX of Streptococcus equisimilis (see AAY06206) and Streptococcus
XX pyogenes. Km values are about 2- to 3-fold lower for UDP sugars,
XX and Vmax values are about 2- to 3-fold higher. The invention
XX provides recombinant vectors containing hyaluronate synthase DNA,
XX especially S. equisimilis hyaluronate synthase DNA (see AAX58841),
XX and prokaryotic or eukaryotic host cells which produce the enzyme
XX and its hyaluronate acid product, particularly a product with
XX modified structure or molecular size. The hyaluronate acid produced
XX this way is purer than that produced by conventional methods.
XX
XX Sequence 972 AA:
SQ
Query Match 11.8%; Score 209; DB 20; Length 972;
Best Local Similarity 21.8%; Pred. No. 1.5e-12;
Matches 96; Conservative 60; Mismatches 146; Indels 138; Gaps 14;
OY 3 PLYSVLICAYNAEKYPAOGLAAYVGGTWRNLDILYDGDSTGTPTAIARHPQDGRIRI 62
DB 440 PLYSVITPYANCANYIQRCVDSALNQTVDLEVCICNDGSTDTTLEVIKLYGNRPVR1 499
OY 63 ISNPNRLGFIASLNLGDLDELASGGGEYIARTDADIASPGWIEKIVGEMERDRIIANG 122
DB 500 MSKPN--GGIASASNAAYSEFAK--GYTIGQSDSDYLEPDAVELCLKEFLKDKTLAC-- 552
OY 123 AMLEVLSENNKSVLAALRNGAIDMKPRHEDIYAVFFPGNPIHNTMIMRS--VIDG 180
DB 553 ---YTTNRNPNPDSILANGYNMPEFSREKLTAMI-----AHFRMFTIRAMHLTDG 603
OY 181 GLRFPAYIHAEDYKFWYAGKIGRLAYYPEALVKYRFHODOTSSK----- 226
DB 604 ---FNKEIENAVDYDFKLSEVGKFKHLNKCYNRVLHGDMTSIKKLGIOKKNHFVVN 660
OY 227 -----YNLQO-----RRTAW----- 236
DB 661 QSLNRQGITYYNDEFDDLDESRRYIFNKTAEQEIEDILDKIKTIONKDAKIAVSIFYP 720
OY 237 -----KIKEIRAGYWK--AAGIAGGACCLNY 261
DB 721 NTLNGLVKKLNMIIEYNKIEYIVLHVDKNHLTPDIKKEILAEYHKNQVNIILNNDISYY 780
OY 262 ---GLKSTAYVALYEKALSGODIGCL-----RLFL---YEVLSLEKYSLTDLDF 306
DB 781 TSNRLIKTAHLSNINKLSQNLNCEYIIFDNHDSLFAVANDSYAY---KKKIDVGNMESA 837
OY 307 LTRDYVRKLEFAAPQYRKILK 326
DB 838 LTHDWIEKINAHBPFFKLIK 857
RESULT 15
AAY43099 standard; Protein; 972 AA.
XX
XX AAY43099;
XX
XX 01-FEB-2000 (first entry)
XX
```

```
DE P. multocida hyaluronate synthase (PmHAS) amino acid sequence.
XX
XX Hyaluronate synthase; PmHAS; hyaluronan; hyaluronate acid; HA; cosmetic;
XX drug delivery; angiogenesis; wound healing; capsule synthesis;
XX fowl cholera; shipping fever.
XX
XX Pasturella multocida.
XX
XX WO951265-A1.
XX
XX 14-OCT-1999.
XX
XX 01-APR-1999; 99WO-US07289.
XX
XX 02-APR-1998; 98US-0080414.
XX
XX 26-OCT-1998; 98US-0178851.
XX
XX (OKLA ) UNIV OKLAHOMA.
XX
XX Deangelis P;
XX
XX WPI: 2000-013032/01.
XX
XX N-PSDB: AA235589.
XX
XX New isolated hyaluronate synthase nucleic acids, used for the
XX production of hyaluronate acid, for developing antibiotics and vaccines
XX and for diagnostic applications
XX
XX Claim 70; Fig 20; 121pp; English.
XX
XX This is the Pasturella multocida hyaluronate synthase (PmHAS) amino acid
XX sequence. Hyaluronate acid (HA) or hyaluronan, is a polysaccharide that
XX serves both structural and recognition roles in higher animals. Bacteria
XX produce extracellular capsules of HA which mimic their host HA and aid
XX escape from a host immune response. The invention includes a vector
XX containing the PmHAS nucleotide sequence which can be used to express
XX PmHAS in a foreign host. The HS nucleic acids can be used for the
XX production of HA. Also, specific changes to the HS coding sequence can
XX result in the production of HA having a modified size distribution or
XX structural configuration and functional properties. The HA products can
XX be used in e.g. drug delivery, angiogenesis and wound healing,
XX stabilisation of recombinant proteins and in cosmetics. The HS nucleic
XX acids can also be used to develop agents to block capsule synthesis by
XX CC pathogens and act as antibiotics. The avirulent P. multocida strains can
XX be used as vaccines for fowl cholera or shipping fever.
XX
XX Sequence 972 AA:
SQ
Query Match 11.8%; Score 209; DB 21; Length 972;
Best Local Similarity 21.8%; Pred. No. 1.5e-12;
Matches 96; Conservative 60; Mismatches 146; Indels 138; Gaps 14;
OY 3 PLYSVLICAYNAEKYPAOGLAAYVGGTWRNLDILYDGDSTGTPTAIARHPQDGRIRI 62
DB 440 PLYSVITPYANCANYIQRCVDSALNQTVDLEVCICNDGSTDTTLEVIKLYGNRPVR1 499
OY 63 ISNPNRLGFIASLNLGDLDELASGGGEYIARTDADIASPGWIEKIVGEMERDRIIANG 122
DB 500 MSKPN--GGIASASNAAYSEFAK--GYTIGQSDSDYLEPDAVELCLKEFLKDKTLAC-- 552
OY 123 AMLEVLSENNKSVLAALRNGAIDMKPRHEDIYAVFFPGNPIHNTMIMRS--VIDG 180
DB 553 ---YTTNRNPNPDSILANGYNMPEFSREKLTAMI-----AHFRMFTIRAMHLTDG 603
OY 181 GLRFPAYIHAEDYKFWYAGKIGRLAYYPEALVKYRFHODOTSSK----- 226
DB 604 ---FNKEIENAVDYDFKLSEVGKFKHLNKCYNRVLHGDMTSIKKLGIOKKNHFVVN 660
OY 227 -----YNLQO-----RRTAW----- 236
DB 661 QSLNRQGITYYNDEFDDLDESRRYIFNKTAEQEIEDILDKIKTIONKDAKIAVSIFYP 720
OY 237 -----KIKEIRAGYWK--AAGIAGGACCLNY 261
```

Db	721	NTLNGLYKLNIIIEYKNIFVIVLHVHDKNHLTPD	IKKEILAFYHKHQNILLNNDISYY	780
Qy	262	---GLKSTAYALYEKALSGODIGCL-----	RUFL---YEXFLSEKYSITDILDF	306
Db	781	TSNRLLITEAHLNSINKLSQNLNCEYIIFDNHDSLF	VKNDSYAY--MKKYDGMNESA	837
Qy	307	LTDVVMRKLFAPQYRKILK	326	
Db	838	LTHDWIEKINAHPPKKLIK	857	

Search completed: December 2, 2002, 11:59:02
Job time : 35.9785 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 12.544 Seconds
(without alignments)
790.458 Million cell updates/sec

Title: US-10-007-267-5

Perfect score: 1764

Sequence: 1 MOPVSVLICAVNAEKYFAQ.....APQYRKILKKMLRPMKYSY 337

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 262574 segs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A.COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B.COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCRTUS.COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfill1.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1764	100.0	337	1	US-08-312-387B-5
2	1764	100.0	337	1	US-08-683-426-5
3	1764	100.0	337	1	US-08-683-426-5
4	1764	100.0	337	2	US-08-878-360-5
5	1764	100.0	337	4	US-09-333-412-5
6	1761	99.8	337	1	US-08-312-387B-12
7	1761	99.8	337	1	US-08-683-426-12
8	1761	99.8	337	1	US-08-683-426-12
9	1761	99.8	337	2	US-08-878-360-12
10	1761	99.8	337	3	US-08-478-140B-5
11	1761	99.8	337	4	US-09-333-412-12
12	1761	99.8	337	4	US-09-338-943-5
13	1049	59.5	348	1	US-08-312-387B-3
14	1049	59.5	348	1	US-08-683-426-3
15	1049	59.5	348	1	US-08-683-426-3
16	1049	59.5	348	2	US-08-878-360-3
17	1049	59.5	348	4	US-09-333-412-3
18	1046	59.3	348	1	US-08-312-387B-11
19	1046	59.3	348	1	US-08-683-426-11
20	1046	59.3	348	1	US-08-683-426-11
21	1046	59.3	348	2	US-08-878-360-11
22	1046	59.3	348	3	US-08-478-140B-3
23	1046	59.3	348	4	US-09-333-412-11
24	1046	59.3	348	4	US-09-338-943-8
25	1046	59.3	348	4	US-09-338-943-8
26	1046	59.3	348	4	US-09-338-943-8
27	209	11.8	965	4	US-09-437-277-3

28	201.5	11.4	702	4	US-09-437-277-1	Sequence 1, Appl1
29	199	11.3	324	1	US-08-597-236-10	Sequence 10, Appl
30	199	11.3	324	1	US-08-746-682A-10	Sequence 10, Appl
31	185.5	10.5	281	4	US-08-961-083-196	Sequence 196, App
32	162	9.2	674	4	US-08-961-083-200	Sequence 200, App
33	152.5	8.6	93	4	US-08-858-207A-521	Sequence 521, App
34	152	8.6	303	4	US-08-961-083-202	Sequence 202, App
35	150	8.5	270	4	US-08-961-083-198	Sequence 198, App
36	138	7.8	358	4	US-09-134-001C-5633	Sequence 5633, Ap
37	131.5	7.5	727	4	US-09-134-001C-4067	Sequence 4067, Ap
38	116.5	6.6	79	4	US-08-961-083-168	Sequence 168, App
39	116	6.6	418	4	US-09-134-001C-4051	Sequence 4051, Ap
40	99.5	5.6	1410	3	US-09-335-409-3	Sequence 3, Appl1
41	99.5	5.6	1410	4	US-09-568-102-3	Sequence 3, Appl1
42	99.5	5.6	1410	4	US-09-567-969-3	Sequence 3, Appl1
43	99.5	5.6	1410	4	US-09-568-480-3	Sequence 3, Appl1
44	99.5	5.6	1410	4	US-09-568-486-3	Sequence 3, Appl1
45	99.5	5.6	1410	4	US-09-568-472-3	Sequence 3, Appl1

ALIGNMENTS

```
RESULT 1
US-08-312-387B-5
; Sequence 5, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gottschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-312-387B-5

Query Match      100.0%; Score 1764; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-178;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 MOPVSVLICAVNAEKYFAQSLAAVVGOTWRNLDILIVDGSIDGTPATARRHEDDGR 60
|||||
DB 1 MOPVSVLICAVNAEKYFAQSLAAVVGOTWRNLDILIVDGSIDGTPATARRHEDDGR 60
|||||
OY 61 RAISNPNLIGFTASINIGDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 120

|||||
Db 61 RIISNPNLGFIA SLNIGDELAKSGGGEYIARTDADDIASPGMIKXIYEMEKDRSIIA 120
QY 121 MGAWLEVLSENNKSVLAAIARNGAIWDKPTRHEDIYVAFPPGNPIHNNMTIMRRSVIDG 180
Db 121 MGAWLEVLSENNKSVLAAIARNGAIWDKPTRHEDIYVAFPPGNPIHNNMTIMRRSVIDG 180
QY 181 GLRFDPAVYHAEDYKFMWEAGKIGRLAYYPEALVYKTRFHODOTSSKYNLQOORTAKIKE 240
Db 181 GLRFDPAVYHAEDYKFMWEAGKIGRLAYYPEALVYKTRFHODOTSSKYNLQOORTAKIKE 240
QY 241 EIRAGYKKAAGIAGVADCLNLYGLKSTAYALYKALSGODIGCLRFLVEYFLSLEKYSI 300
Db 241 EIRAGYKKAAGIAGVADCLNLYGLKSTAYALYKALSGODIGCLRFLVEYFLSLEKYSI 300
QY 301 TDLLDFLTDVRMKLFAAPQYRKILKMLRPMKYRSY 337
Db 301 TDLLDFLTDVRMKLFAAPQYRKILKMLRPMKYRSY 337

RESULT 2

US-08-683-426-5
; Sequence 5, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,426
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-0958
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ. ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-426-5

Query Match 100.0%; Score 1764; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-178;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 MOPVLSVLCAYNAEKYFAQSILAAVVGQGTWRNDILIVDDGSTDGTPTAIRHRFOEDGRI 60
Db 1 MOPVLSVLCAYNAEKYFAQSILAAVVGQGTWRNDILIVDDGSTDGTPTAIRHRFOEDGRI 60

QY 61 RIISNPNLGFIA SLNIGDELAKSGGGEYIARTDADDIASPGMIKXIYEMEKDRSIIA 120
Db 61 RIISNPNLGFIA SLNIGDELAKSGGGEYIARTDADDIASPGMIKXIYEMEKDRSIIA 120
QY 121 MGAWLEVLSENNKSVLAAIARNGAIWDKPTRHEDIYVAFPPGNPIHNNMTIMRRSVIDG 180
Db 121 MGAWLEVLSENNKSVLAAIARNGAIWDKPTRHEDIYVAFPPGNPIHNNMTIMRRSVIDG 180
QY 181 GLRFDPAVYHAEDYKFMWEAGKIGRLAYYPEALVYKTRFHODOTSSKYNLQOORTAKIKE 240
Db 181 GLRFDPAVYHAEDYKFMWEAGKIGRLAYYPEALVYKTRFHODOTSSKYNLQOORTAKIKE 240
QY 241 EIRAGYKKAAGIAGVADCLNLYGLKSTAYALYKALSGODIGCLRFLVEYFLSLEKYSI 300
Db 241 EIRAGYKKAAGIAGVADCLNLYGLKSTAYALYKALSGODIGCLRFLVEYFLSLEKYSI 300
QY 301 TDLLDFLTDVRMKLFAAPQYRKILKMLRPMKYRSY 337
Db 301 TDLLDFLTDVRMKLFAAPQYRKILKMLRPMKYRSY 337

RESULT 3

US-08-683-458-5
; Sequence 5, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ. ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
; US-08-683-458-5

Query Match 100.0%; Score 1764; DB 1; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-178;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Caps 0;

QY 1 MOPVLSVLCAYNAEKYFAQSILAAVVGQGTWRNDILIVDDGSTDGTPTAIRHRFOEDGRI 60
Db 1 MOPVLSVLCAYNAEKYFAQSILAAVVGQGTWRNDILIVDDGSTDGTPTAIRHRFOEDGRI 60

QY 61 RIISNPNLGIASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 120
DB 61 RIISNPNLGIASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 120
QY 121 MGAMLEVLSEENNSVLAIAIRNGAIIWDKPTRHEDIYAVPFPGNPIHNTMIIMRSVIDG 180
DB 121 MGAMLEVLSEENNSVLAIAIRNGAIIWDKPTRHEDIYAVPFPGNPIHNTMIIMRSVIDG 180
QY 181 GLRPDPATIAHEDYKFWYEAGKLGRLAYPPALVYKRRHQDQTSKYNLQORRTAMKIKE 240
DB 181 GLRPDPATIAHEDYKFWYEAGKLGRLAYPPALVYKRRHQDQTSKYNLQORRTAMKIKE 240
QY 241 EIRAGYKAAAGIAGADCLNGLKSTAYALYERKALSGODIGCLRFLYEYFLSLEKYSL 300
DB 241 EIRAGYKAAAGIAGADCLNGLKSTAYALYERKALSGODIGCLRFLYEYFLSLEKYSL 300
QY 301 TDLDLFTDRVWRKLFAPAORYRKILKMLRPWKYRSY 337
DB 301 TDLDLFTDRVWRKLFAPAORYRKILKMLRPWKYRSY 337

RESULT 4
US-08-878-360-5

Sequence 5, Application US/08878360
Patent No. 5945322
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/683,426
FILING DATE:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-360-5

Query Match 100.0%; Score 1764; DB 2; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-178;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLVSVLICAYNAEKYFAQSLAAVVGQWNRNLDILYVDSSTGCTPAIAIARHFOEDGRI 60
DB 1 MQLVSVLICAYNAEKYFAQSLAAVVGQWNRNLDILYVDSSTGCTPAIAIARHFOEDGRI 60
QY 61 RIISNPNLGIASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 120
DB 61 RIISNPNLGIASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSIIA 120
QY 121 MGAMLEVLSEENNSVLAIAIRNGAIIWDKPTRHEDIYAVPFPGNPIHNTMIIMRSVIDG 180
DB 121 MGAMLEVLSEENNSVLAIAIRNGAIIWDKPTRHEDIYAVPFPGNPIHNTMIIMRSVIDG 180
QY 181 GLRPDPATIAHEDYKFWYEAGKLGRLAYPPALVYKRRHQDQTSKYNLQORRTAMKIKE 240
DB 181 GLRPDPATIAHEDYKFWYEAGKLGRLAYPPALVYKRRHQDQTSKYNLQORRTAMKIKE 240
QY 241 EIRAGYKAAAGIAGADCLNGLKSTAYALYERKALSGODIGCLRFLYEYFLSLEKYSL 300
DB 241 EIRAGYKAAAGIAGADCLNGLKSTAYALYERKALSGODIGCLRFLYEYFLSLEKYSL 300
QY 301 TDLDLFTDRVWRKLFAPAORYRKILKMLRPWKYRSY 337
DB 301 TDLDLFTDRVWRKLFAPAORYRKILKMLRPWKYRSY 337

RESULT 5
US-09-333-412-5

Sequence 5, Application US/09333412
Patent No. 642382
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESSES:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-09-333-412-5

Query Match 100.0%; Score 1764; DB 4; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.2e-178;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;


```
Db      1  LQPLVSVLLICAVNAEKYFPAQSLAAVVGQTMRLDILIVDGGSTDTGPAIARHFQEDGRI 60
QY      61  RIISNPRNIGFTIASLNLIGIDELAKSGGGEYIARTDADDIASFGMIETKITYGEMEKDSITIA 120
Db      61  RIISNPRNIGFTIASLNLIGIDELAKSGGGEYIARTDADDIASFGMIETKITYGEMEKDSITIA 120
QY      121  MGAMLEVLSEENKNSVLAIAIRNGAIWMDKPTRHEDIIVAVFPFGNP1HNNTMTIMRSVIDG 180
Db      121  MGAMLEVLSEENKNSVLAIAIRNGAIWMDKPTRHEDIIVAVFPFGNP1HNNTMTIMRSVIDG 180
QY      181  GLRFPAYIHAEDYKFWYAGKLGRLAYYPEALVYKRFHQDOTSSKYNLQORRTAWKIKE 240
Db      181  GLRFPAYIHAEDYKFWYAGKLGRLAYYPEALVYKRFHQDOTSSKYNLQORRTAWKIKE 240
QY      241  EIRAGYWKAAAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRLFLYEYFLSLEKYSL 300
Db      241  EIRAGYWKAAAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRLFLYEYFLSLEKYSL 300
QY      301  TDLLDFLTDVRMKLLEAPQYRKILKMLRPWKYRSY 337
Db      301  TDLLDFLTDVRMKLLEAPQYRKILKMLRPWKYRSY 337
```

RESULT 8

```
US-08-683-458-12
; Sequence 12, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095A
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-683-458-12
```

```
Query Match          99.8%; Score 1761; DB 1; Length 337;
Best Local Similarity 99.7%; Pred. No. 2,4e-178;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;
```

```
QY      1  MOPVSVLLICAVNAEKYFPAQSLAAVVGQTMRLDILIVDGGSTDTGPAIARHFQEDGRI 60
Db      1  LQPLVSVLLICAVNAEKYFPAQSLAAVVGQTMRLDILIVDGGSTDTGPAIARHFQEDGRI 60
QY      61  RIISNPRNIGFTIASLNLIGIDELAKSGGGEYIARTDADDIASFGMIETKITYGEMEKDSITIA 120
Db      61  RIISNPRNIGFTIASLNLIGIDELAKSGGGEYIARTDADDIASFGMIETKITYGEMEKDSITIA 120
QY      121  MGAMLEVLSEENKNSVLAIAIRNGAIWMDKPTRHEDIIVAVFPFGNP1HNNTMTIMRSVIDG 180
Db      121  MGAMLEVLSEENKNSVLAIAIRNGAIWMDKPTRHEDIIVAVFPFGNP1HNNTMTIMRSVIDG 180
QY      181  GLRFPAYIHAEDYKFWYAGKLGRLAYYPEALVYKRFHQDOTSSKYNLQORRTAWKIKE 240
Db      181  GLRFPAYIHAEDYKFWYAGKLGRLAYYPEALVYKRFHQDOTSSKYNLQORRTAWKIKE 240
QY      241  EIRAGYWKAAAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRLFLYEYFLSLEKYSL 300
Db      241  EIRAGYWKAAAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRLFLYEYFLSLEKYSL 300
QY      301  TDLLDFLTDVRMKLLEAPQYRKILKMLRPWKYRSY 337
Db      301  TDLLDFLTDVRMKLLEAPQYRKILKMLRPWKYRSY 337
```

RESULT 9

```
US-08-878-360-12
; Sequence 12, Application US/08878360
; Patent No. 5945322
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; COMPUTER: IBM PC compatible
; OPERATING SYSTEM: PC-DOS/MS-DOS
; SOFTWARE: PatentIn Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/878,360
; FILING DATE: 18-JUN-1997
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/683,426
; FILING DATE:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095B
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; INFORMATION FOR SEQ ID NO: 12:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-878-360-12
```

```
Query Match          99.8%; Score 1761; DB 2; Length 337;
```

Best Local Similarity 99.7%; Pred. No. 2.4e-178;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLLICAVAAEKYFQSLAAVVGOTWRNDDIIVDDGSDGTGPALARRHQEDGR1 60
Db 1 LOPVSVLLICAVAAEKYFQSLAAVVGOTWRNDDIIVDDGSDGTGPALARRHQEDGR1 60
QY 61 RIISNPNRNGFIASLNIIGDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSSIIA 120
Db 61 RIISNPNRNGFIASLNIIGDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSSIIA 120
QY 121 MGAWEVLSEENKSVLAAILARNGAIWDRPTREDIIVAVFPFGNP1HNNTMTIMRSVIDG 180
Db 121 MGAWEVLSEENKSVLAAILARNGAIWDRPTREDIIVAVFPFGNP1HNNTMTIMRSVIDG 180
QY 181 GLRFDPAYTHAEDYKFWYEGKIGRLAYYPEALVKYRFHODOTSRYN1QOORTANKIKE 240
Db 181 GLRFDPAYTHAEDYKFWYEGKIGRLAYYPEALVKYRFHODOTSRYN1QOORTANKIKE 240
QY 241 EIRAGYWKAGIAGVADCLNYGLKSTAVYALKSGODIGCLRFLYEYFLSEKYSYL 300
Db 241 EIRAGYWKAGIAGVADCLNYGLKSTAVYALKSGODIGCLRFLYEYFLSEKYSYL 300
QY 301 TDLLDFLTRVVRMKLFPAAPQYRKILKKMLRPWKYRSY 337
Db 301 TDLLDFLTRVVRMKLFPAAPQYRKILKKMLRPWKYRSY 337

RESULT 10

US-08-478-140B-5
; Sequence 5, Application US/08478140B
; Patent No. 6127153

GENERAL INFORMATION:

APPLICANT: JOHNSON, KARL F.
APPLICANT: BUCZALA, STEPHANIE L.
TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
NUMBER OF SEQUENCES: 8
CORRESPONDENCE ADDRESS:
ADDRESSEE: Pennie & Edmonds
STREET: 1155 Avenue of the Americas
CITY: New York
STATE: New York
COUNTRY: USA
ZIP: 10036-2711

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/478.140B
FILING DATE: 07-JUN-1995
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Laura A. Coruzzi

REGISTRATION NUMBER: 30,742
REFERENCE/DOCKET NUMBER: 7188-017
TELECOMMUNICATION INFORMATION:
TELEPHONE: (212) 790-9090
TELEFAX: (212) 869-9741/8864
TELEX: 66141 PENNTE

INFORMATION FOR SEQ ID NO: 5:

SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-478-140B-5

Query Match 99.8%; Score 1761; DB 3; Length 337;
Best Local Similarity 99.7%; Pred. No. 2.4e-178;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOPVSVLLICAVAAEKYFQSLAAVVGOTWRNDDIIVDDGSDGTGPALARRHQEDGR1 60
Db 1 LOPVSVLLICAVAAEKYFQSLAAVVGOTWRNDDIIVDDGSDGTGPALARRHQEDGR1 60
QY 61 RIISNPNRNGFIASLNIIGDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSSIIA 120
Db 61 RIISNPNRNGFIASLNIIGDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSSIIA 120
QY 121 MGAWEVLSEENKSVLAAILARNGAIWDRPTREDIIVAVFPFGNP1HNNTMTIMRSVIDG 180
Db 121 MGAWEVLSEENKSVLAAILARNGAIWDRPTREDIIVAVFPFGNP1HNNTMTIMRSVIDG 180
QY 181 GLRFDPAYTHAEDYKFWYEGKIGRLAYYPEALVKYRFHODOTSRYN1QOORTANKIKE 240
Db 181 GLRFDPAYTHAEDYKFWYEGKIGRLAYYPEALVKYRFHODOTSRYN1QOORTANKIKE 240
QY 241 EIRAGYWKAGIAGVADCLNYGLKSTAVYALKSGODIGCLRFLYEYFLSEKYSYL 300
Db 241 EIRAGYWKAGIAGVADCLNYGLKSTAVYALKSGODIGCLRFLYEYFLSEKYSYL 300
QY 301 TDLLDFLTRVVRMKLFPAAPQYRKILKKMLRPWKYRSY 337
Db 301 TDLLDFLTRVVRMKLFPAAPQYRKILKKMLRPWKYRSY 337

RESULT 11

US-09-333-412-12
; Sequence 12, Application US/09333412
; Patent No. 6342382

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-09-333-412-12

Query Match 99.8%; Score 1761; DB 4; Length 337;
Best Local Similarity 99.7%; Pred. No. 2.4e-178;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLPLSVLICAVNAEKYFQSQSLAAVVGQWRNLDLIYDDGSTDGTGPATARRHFOEDGR1 60
:|||||
DB 1 LQPLSVLICAVNAEKYFQSQSLAAVVGQWRNLDLIYDDGSTDGTGPATARRHFOEDGR1 60

QY 61 RIISPNRNGFTASINIGDELAKSGGGEYIARTDADDIASGWIETKYEKEXRSTIA 120
:|||||
DB 61 RIISPNRNGFTASINIGDELAKSGGGEYIARTDADDIASGWIETKYEKEXRSTIA 120

QY 121 MGAWEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIIVAVFPNGNPIHNTMTIMRSVIDG 180
:|||||
DB 121 MGAWEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIIVAVFPNGNPIHNTMTIMRSVIDG 180

QY 181 GLRFPAYTHADYKFWYBAGKLGRLAYYPEALVKYRFHQDTSKYNLQORRTAMKIKE 240
:|||||
DB 181 GLRFPAYTHADYKFWYBAGKLGRLAYYPEALVKYRFHQDTSKYNLQORRTAMKIKE 240

QY 241 EIRAGYWRKAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRFLYEYFLSEKYSL 300
:|||||
DB 241 EIRAGYWRKAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRFLYEYFLSEKYSL 300

QY 301 TDLDLFLTRVNRKLFAPQYRKILKMLRPMKYRSY 337
:|||||
DB 301 TDLDLFLTRVNRKLFAPQYRKILKMLRPMKYRSY 337

RESULT 12
US-09-338-943-5
; Sequence 5, Application US/09338943
; Patent No. 6379933
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, KARL F.
; APPLICANT: ROTH, STEPHEN
; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE, A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/338,943
; FILING DATE:
; CLASSIFICATION:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: US 08/478,140
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEFAX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO. 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 337 amino acids

TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-09-338-943-5

Query Match 99.8%; Score 1761; DB 4; Length 337;
Best Local Similarity 99.7%; Pred. No. 2.4e-178;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLPLSVLICAVNAEKYFQSQSLAAVVGQWRNLDLIYDDGSTDGTGPATARRHFOEDGR1 60
:|||||
DB 1 LQPLSVLICAVNAEKYFQSQSLAAVVGQWRNLDLIYDDGSTDGTGPATARRHFOEDGR1 60

QY 61 RIISPNRNGFTASINIGDELAKSGGGEYIARTDADDIASGWIETKYEKEXRSTIA 120
:|||||
DB 61 RIISPNRNGFTASINIGDELAKSGGGEYIARTDADDIASGWIETKYEKEXRSTIA 120

QY 121 MGAWEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIIVAVFPNGNPIHNTMTIMRSVIDG 180
:|||||
DB 121 MGAWEVLSEENKSVLAIAIRNGAIWDPKPTRHEDIIVAVFPNGNPIHNTMTIMRSVIDG 180

QY 181 GLRFPAYTHADYKFWYBAGKLGRLAYYPEALVKYRFHQDTSKYNLQORRTAMKIKE 240
:|||||
DB 181 GLRFPAYTHADYKFWYBAGKLGRLAYYPEALVKYRFHQDTSKYNLQORRTAMKIKE 240

QY 241 EIRAGYWRKAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRFLYEYFLSEKYSL 300
:|||||
DB 241 EIRAGYWRKAGIAGVADCLNTYGLKSTAYALYEKALSGODIGCLRFLYEYFLSEKYSL 300

QY 301 TDLDLFLTRVNRKLFAPQYRKILKMLRPMKYRSY 337
:|||||
DB 301 TDLDLFLTRVNRKLFAPQYRKILKMLRPMKYRSY 337

RESULT 13
US-08-312-387B-3
; Sequence 3, Application US/08312387B
; Patent No. 5545553
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauber & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/312,387B
; FILING DATE: July 7, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REGISTRATION NUMBER: 26,742
; REFERENCE/DOCKET NUMBER: 600-1-095
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEFAX: 133521
; INFORMATION FOR SEQ ID NO. 3:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 348 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein

US-08-312-387B-3

Query Match 59.5%; Score 1049; DB 1; Length 348;

Best Local Similarity 64.5%; Pred. No. 7.3e-103;

Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

```
QY 1 MOPVSVLICAVNAEKYFQSLAAVVGOTWRNLDILYDDGSTDGTPAARHFOEDGRI 60
    |||||
DB 1 MOPVSVLICAVNAEKYFQSLAAVVGOTWRNLDILYDDGSTDGTLATAKDEQKRSRI 60
QY 61 RIISNPNRGFLASLTNGIDELAKS--GGGEYIARTDADDIASPGITEKIVGEMEDRSI 118
    :|:::|
DB 61 KILAAQNSGLPLSLNIGIDELAKSGGGGEYIARTDADDIASPGITEKIVGEMEDRSI 120
QY 119 IIMGANLEYLESENNKSVLAATARNGAINDKPTRHEDIYAVPEFGPNINNTMIMRSYI 178
    |||||
DB 121 IIMGANLEYLESENNKSVLAATARNGAINDKPTRHEDIYAVPEFGPNINNTMIMRSYI 180
QY 121 IIMGANLEYLESENNKSVLAATARNGAINDKPTRHEDIYAVPEFGPNINNTMIMRSYI 180
QY 179 DGLRDPAYIHAEDYKFWYKGLRLAYYPEALVKYRFHODOTSSKYNLQORRTAMKI 238
    |||||
DB 181 DGLRDPAYIHAEDYKFWYKGLRLAYYPEALVKYRFHODOTSSKYNLQORRTAMKI 240
QY 239 KEIRAGYKKAAGIAGADCLNGLIKSTAYALYKALSGODIGCLRLYLEYFLSLEY 298
    :|:::|
DB 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAYELPEKDLPEDEFERARFLYOCFKRTDTP 300
QY 299 SLTDLDPLTDYMRKLFAPQYRKILKKMLR 330
    |||
DB 301 PSGAMLDFAADGMRRLFTLRQYFGILYRLIK 332
```

RESULT 14

US-08-683-426-3

Sequence 3, Application US/08683426

Patent No. 5705367

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/683,426

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/312,387

FILING DATE: September 26, 1994

CLASSIFICATION: 536

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-0958

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 348 amino acids

TYPE: amino acid

TOPOLOGY: linear

MOLECULE TYPE: protein

Query Match 59.5%; Score 1049; DB 1; Length 348;

Best Local Similarity 64.5%; Pred. No. 7.3e-103;

Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

```
QY 1 MOPVSVLICAVNAEKYFQSLAAVVGOTWRNLDILYDDGSTDGTPAARHFOEDGRI 60
    |||||
DB 1 MOPVSVLICAVNAEKYFQSLAAVVGOTWRNLDILYDDGSTDGTLATAKDEQKRSRI 60
QY 61 RIISNPNRGFLASLTNGIDELAKS--GGGEYIARTDADDIASPGITEKIVGEMEDRSI 118
    :|:::|
DB 61 KILAAQNSGLPLSLNIGIDELAKSGGGGEYIARTDADDIASPGITEKIVGEMEDRSI 120
QY 119 IIMGANLEYLESENNKSVLAATARNGAINDKPTRHEDIYAVPEFGPNINNTMIMRSYI 178
    |||||
DB 121 IIMGANLEYLESENNKSVLAATARNGAINDKPTRHEDIYAVPEFGPNINNTMIMRSYI 180
QY 121 IIMGANLEYLESENNKSVLAATARNGAINDKPTRHEDIYAVPEFGPNINNTMIMRSYI 180
QY 179 DGLRDPAYIHAEDYKFWYKGLRLAYYPEALVKYRFHODOTSSKYNLQORRTAMKI 238
    |||||
DB 181 DGLRDPAYIHAEDYKFWYKGLRLAYYPEALVKYRFHODOTSSKYNLQORRTAMKI 240
QY 239 KEIRAGYKKAAGIAGADCLNGLIKSTAYALYKALSGODIGCLRLYLEYFLSLEY 298
    :|:::|
DB 241 OKTARNDFLOSMGFKTRFDSLEYRQTKAAYELPEKDLPEDEFERARFLYOCFKRTDTP 300
QY 299 SLTDLDPLTDYMRKLFAPQYRKILKKMLR 330
    |||
DB 301 PSGAMLDFAADGMRRLFTLRQYFGILYRLIK 332
```

RESULT 15

US-08-683-458-3

Sequence 3, Application US/08683458

Patent No. 5798233

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/683,458

FILING DATE:

CLASSIFICATION: 435

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 08/312,387

FILING DATE: September 26, 1994

CLASSIFICATION: 435

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-0958

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 3:

SEQUENCE CHARACTERISTICS:

LENGTH: 348 amino acids

TYPE: amino acid

; TOPOLOGY: linear
; MOLECULE TYPE: protein
us-08-683-458-3

Query Match 59.5%; Score 1049; DB 1; Length 348;

Best Local Similarity 64.5%; Pred. No. 7.3e-103;
Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

QY 1 MOPVSVLICAYNAEKYFAOSLAAYVGGTWRNLDIYDDGSTDGTPAIRHFOEDGRI 60
Db 1 MOPVSVLICAYNAEKYFAOSLAAYVGGTWRNLDIYDDGSTDGTPAIRHFOEDGRI 60
QY 61 RIISNPNRLGFTASINIGLDELAKS--GGGEYIARTDADDIASPGWIEKIYGEEMEKDRSI 118
Db 61 KILAQAQNSGLIPSLINIGLDELAKSGGGGEYIARTDADDIASPGWIEKIYGEEMEKDRSI 120
QY 119 IAMGAMLEVLSEENKSVLAIAIRNGAIIWDKPTRHEDIYAVFPFGNPPIHNTMTIMRSYI 178
Db 121 IAMGAMLEVLSEENKSVLAIAIRNGAIIWDKPTRHEDIYAVFPFGNPPIHNTMTIMRSYI 180
QY 179 DGGLRFPDPAVTHAEDYKFWYEAGKLGRLAYYPEALVYKRFHODTSSKYNLQORRTAMKI 238
Db 181 DGGLRFPDPAVTHAEDYKFWYEAGKLGRLAYYPEALVYKRFHODTSSKYNLQORRTAMKI 240
QY 239 KEETIRAGYWKAAAGIAGVADCLNYGLKSTAYALYEKALSGODIGCLRFLYEYFLSLEKY 298
Db 241 OKTARNDFLOSMGFEKTRFDSLEYRQTAAAYELPEKDLPEEDFERARRLFVQCFFKRTDTP 300
QY 299 SITDLIDFLTDRAWMKLPAAPQYRKILKKMLR 330
Db 301 PSGAWLIDFAADGRMRRLFTLRQYFGILYRLIK 332

Search completed: December 2, 2002, 12:03:15
Job time : 13.544 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 12:00:15 ; Search time 6.75447 Seconds
(without alignments)
794.504 Million cell updates/sec

Title: US-10-007-267-5
1764
Sequence: 1 MQLPLVSLICAYNAEYFAQ.....APQYRKILKMLRPWKYSY 337

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Published_Applications_AA:*
1: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB.pep:*
2: /cgn2_6/ptodata/1/pubpaa/PCR_NEM_PUB.pep:*
3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB.pep:*
4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*
5: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*
7: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*
9: /cgn2_6/ptodata/1/pubpaa/US09_NEM_PUB.pep:*
10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*
11: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB.pep:*
12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*
13: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB.pep:*
14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1764	100.0	337	12	US-10-007-267-5
2	1761	99.8	337	12	US-10-007-267-12
3	1049	59.5	348	12	US-10-007-267-3
4	1046	59.3	348	12	US-10-007-267-11
5	218	12.4	150	10	US-09-924-358-29
6	209	11.8	972	9	US-09-879-959-10
7	208	11.8	332	10	US-09-767-041-22
8	207.5	11.8	278	10	US-09-767-041-36
9	194	11.0	297	10	US-09-816-028A-31
10	188	10.7	322	10	US-09-767-041-35
11	187.5	10.6	322	10	US-09-767-041-34
12	187	10.6	301	10	US-09-816-028A-27
13	185.5	10.5	120	10	US-09-767-041-51
14	185.5	10.5	270	10	US-09-816-028A-39
15	185.5	10.5	281	10	US-09-765-272-196
16	185.5	10.5	332	10	US-09-767-041-21
17	176.5	10.0	313	10	US-09-900-038A-1
18	174	9.9	210	10	US-09-767-041-47
19	162	9.2	303	10	US-09-816-028A-29

20	162	9.2	674	10	US-09-765-272-200	Sequence 200, App
21	152	8.6	120	10	US-09-767-041-52	Sequence 52, Appl
22	152	8.6	303	10	US-09-765-272-202	Sequence 202, App
23	152	8.6	706	10	US-09-815-242-4950	Sequence 4950, App
24	152	8.6	715	10	US-09-815-242-10511	Sequence 10511, App
25	150	8.5	270	10	US-09-765-272-198	Sequence 198, App
26	147.5	8.4	187	9	US-09-973-457-4	Sequence 4, Appl1
27	147.5	8.4	187	10	US-09-815-028-7	Sequence 7, Appl1
28	147.5	8.4	187	12	US-10-074-527-4	Sequence 4, Appl1
29	145	8.2	389	10	US-09-816-028A-34	Sequence 34, Appl1
30	127	7.2	256	10	US-09-925-301-883	Sequence 883, App
31	123	7.0	608	10	US-09-924-358-8	Sequence 8, Appl1
32	116.5	6.6	79	10	US-09-765-272-168	Sequence 168, App
33	116.5	6.6	269	10	US-09-767-041-41	Sequence 41, Appl1
34	108	6.1	418	10	US-09-816-028A-33	Sequence 33, Appl1
35	105	6.0	358	10	US-09-815-242-5714	Sequence 5714, App
36	105	6.0	573	10	US-09-815-242-12474	Sequence 12474, A
37	103	5.8	612	12	US-10-001-851-25	Sequence 25, Appl1
38	98	5.6	324	9	US-09-981-876-181	Sequence 181, App
39	95	5.4	306	10	US-09-912-020-275	Sequence 275, App
40	95	5.4	639	10	US-09-789-417-2	Sequence 2, Appl1
41	95	5.4	639	10	US-09-792-451-2	Sequence 2, Appl1
42	95	5.4	639	12	US-10-052-586-364	Sequence 364, App
43	94.5	5.4	626	12	US-10-001-851-27	Sequence 27, Appl1
44	93	5.3	844	9	US-10-108-605-185	Sequence 185, App
45	92	5.2	963	10	US-09-801-368-74	Sequence 74, Appl1

ALIGNMENTS

RESULT 1
US-10-007-267-5
Sequence 5, Application US/10007267
Patent No. US20020127682A1
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION/DOCKET NUMBER: 26,742
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid

TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 5:
US-10-007-267-5

Query Match 100.0%; Score 1764; DB 12; Length 337;
Best Local Similarity 100.0%; Pred. No. 1.7e-166;
Matches 337; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLVSVLICAVNAEKYFAQSLAAVVGQTFWRNDILIVDGSYDGTGPALARRHQEDGR 60
DB 1 MQLVSVLICAVNAEKYFAQSLAAVVGQTFWRNDILIVDGSYDGTGPALARRHQEDGR 60
QY 61 RIISNPNRLGFIASLNGIDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSIIA 120
DB 61 RIISNPNRLGFIASLNGIDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSIIA 120
QY 121 MGAWEVLSSENNKSVLAARNGAIWDKPTRHEDIYAVFPFGNPJHNTMIMRRSVIDG 180
DB 121 MGAWEVLSSENNKSVLAARNGAIWDKPTRHEDIYAVFPFGNPJHNTMIMRRSVIDG 180
QY 181 GLRFPDPAIHAEDYKFWYEGKLGRLAYYPEALVKYRFHODQTSKYNLQOORTANKIKE 240
DB 181 GLRFPDPAIHAEDYKFWYEGKLGRLAYYPEALVKYRFHODQTSKYNLQOORTANKIKE 240
QY 241 EIRAGYWKAAAGIAGVADCLNGLKSTAYALYEKALSQDIGCLRFLFYEFLSLEKYS 300
DB 241 EIRAGYWKAAAGIAGVADCLNGLKSTAYALYEKALSQDIGCLRFLFYEFLSLEKYS 300
QY 301 TDLLDFLTRVMRKLPAAPQYRKILKKMLRPWKYRSY 337
DB 301 TDLLDFLTRVMRKLPAAPQYRKILKKMLRPWKYRSY 337

RESULT 2
US-10-007-267-12

Sequence 12, Application US/10007267
Patent No. US20020127682A1

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
OLIGOSACCHARIDES, AND GENES ENCODING THEM

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,267

FILING DATE: 03-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/333,412

FILING DATE: 15-Jun-1999

APPLICATION NUMBER: 08/312,387

FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-095

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 12:

SEQUENCE CHARACTERISTICS:
LENGTH: 337 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 12:
US-10-007-267-12

Query Match 99.8%; Score 1761; DB 12; Length 337;
Best Local Similarity 99.7%; Pred. No. 3.4e-166;
Matches 336; Conservative 1; Mismatches 0; Indels 0; Gaps 0;

QY 1 MQLVSVLICAVNAEKYFAQSLAAVVGQTFWRNDILIVDGSYDGTGPALARRHQEDGR 60
DB 1 MQLVSVLICAVNAEKYFAQSLAAVVGQTFWRNDILIVDGSYDGTGPALARRHQEDGR 60
QY 61 RIISNPNRLGFIASLNGIDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSIIA 120
DB 61 RIISNPNRLGFIASLNGIDELAKSGGGEYIARTDADDIASPGMIKIKYEMEKDSIIA 120
QY 121 MGAWEVLSSENNKSVLAARNGAIWDKPTRHEDIYAVFPFGNPJHNTMIMRRSVIDG 180
DB 121 MGAWEVLSSENNKSVLAARNGAIWDKPTRHEDIYAVFPFGNPJHNTMIMRRSVIDG 180
QY 181 GLRFPDPAIHAEDYKFWYEGKLGRLAYYPEALVKYRFHODQTSKYNLQOORTANKIKE 240
DB 181 GLRFPDPAIHAEDYKFWYEGKLGRLAYYPEALVKYRFHODQTSKYNLQOORTANKIKE 240
QY 241 EIRAGYWKAAAGIAGVADCLNGLKSTAYALYEKALSQDIGCLRFLFYEFLSLEKYS 300
DB 241 EIRAGYWKAAAGIAGVADCLNGLKSTAYALYEKALSQDIGCLRFLFYEFLSLEKYS 300
QY 301 TDLLDFLTRVMRKLPAAPQYRKILKKMLRPWKYRSY 337
DB 301 TDLLDFLTRVMRKLPAAPQYRKILKKMLRPWKYRSY 337

RESULT 3
US-10-007-267-3

Sequence 3, Application US/10007267
Patent No. US20020127682A1

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
OLIGOSACCHARIDES, AND GENES ENCODING THEM

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESS:

ADDRESSEE: Klauber & Jackson

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,267

FILING DATE: 03-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/333,412

FILING DATE: 15-Jun-1999

APPLICATION NUMBER: 08/312,387

FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-095

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684
TELEPHONE: 201 487-5800
INFORMATION FOR SEQ ID NO: 3:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 3:
US-10-007-267-3

Query Match 59.5%; Score 1049; DB 12; Length 348;
Best Local Similarity 64.5%; Pred. No. 6,8e-96;
Matches 214; Conservative 33; Mismatches 83; Indels 2; Gaps 1;

QY 1 MOPVSVLICAVNAEKYFQSLAAVVGOTWRNDILIVDGSSTGTPATARHPQEDGR 60
DB 1 MOPVSVLICAVNAEKYFQSLAAVVGOTWRNDILIVDGSSTGTPATARHPQEDGR 60
QY 61 RIISNPNLGFIALNIGDELAKS--GGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 118
DB 61 KILAQONSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
QY 119 IANGWLEVLSEENKSVLAATRNCAIMDKPRHEDIYAVFPFGNPINHTMIMRSYI 178
DB 121 IANGWLEVLSEENKSVLAATRNCAIMDKPRHEDIYAVFPFGNPINHTMIMRSYI 180
QY 179 DGLRFPDAVIAHEDYKFWYEAQKLGRLAYYPALVKYRPHODTSKYNLQORRTAWI 238
DB 181 DGLRFPDAVIAHEDYKFWYEAQKLGRLAYYPALVKYRPHODTSKYNLQORRTAWI 240
QY 239 KEIRIRAGYKAKAGIAGADCLNGLKSTAYALYERKALSGODIGCLRFLYEFYLSLEY 298
DB 241 QKTAARNDFLOSNGKFRFDSLEYRQTKAAAYELPEKDLPEDEPRARRFLYQCFKRTDTP 300
QY 299 SLTDLDFLTDVYRKLFAPAYRKILKKMLR 330
DB 301 PSGAWLDFADGRMRRLFTLRYFGILYRLIK 332

RESULT 4
US-10-007-267-11
Sequence 11, Application US/10007267
Patent No. US20020127682A1
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSTRANSFERASES FOR BIOSYNTHESIS OF OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION NUMBER: US/10/007,267
FILING DATE: 03-Dec-2001
CLASSIFICATION: <Unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742

REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
INFORMATION FOR SEQ ID NO: 11:
SEQUENCE CHARACTERISTICS:
LENGTH: 348 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 11:
US-10-007-267-11

Query Match 59.3%; Score 1046; DB 12; Length 348;
Best Local Similarity 64.2%; Pred. No. 1.4e-95;
Matches 213; Conservative 34; Mismatches 83; Indels 2; Gaps 1;

QY 1 MOPVSVLICAVNAEKYFQSLAAVVGOTWRNDILIVDGSSTGTPATARHPQEDGR 60
DB 1 MOPVSVLICAVNAEKYFQSLAAVVGOTWRNDILIVDGSSTGTPATARHPQEDGR 60
QY 61 RIISNPNLGFIALNIGDELAKS--GGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 118
DB 61 KILAQONSGLIPSLNIGDELAKSGGGGEYIARTDADDIASPGWIEKIVGEMEKDRSI 120
QY 119 IANGWLEVLSEENKSVLAATRNCAIMDKPRHEDIYAVFPFGNPINHTMIMRSYI 178
DB 121 IANGWLEVLSEENKSVLAATRNCAIMDKPRHEDIYAVFPFGNPINHTMIMRSYI 180
QY 179 DGLRFPDAVIAHEDYKFWYEAQKLGRLAYYPALVKYRPHODTSKYNLQORRTAWI 238
DB 181 DGLRFPDAVIAHEDYKFWYEAQKLGRLAYYPALVKYRPHODTSKYNLQORRTAWI 240
QY 239 KEIRIRAGYKAKAGIAGADCLNGLKSTAYALYERKALSGODIGCLRFLYEFYLSLEY 298
DB 241 QKTAARNDFLOSNGKFRFDSLEYRQTKAAAYELPEKDLPEDEPRARRFLYQCFKRTDTP 300
QY 299 SLTDLDFLTDVYRKLFAPAYRKILKKMLR 330
DB 301 PSGAWLDFADGRMRRLFTLRYFGILYRLIK 332

RESULT 5
US-09-924-358-29
Sequence 29, Application US/09924358
Patent No. US20020107376A1
GENERAL INFORMATION:
APPLICANT: Millennium Pharmaceuticals, Inc.
APPLICANT: Meyers, Rachel
TITLE OF INVENTION: 26199, 33530, 33949, 47148, 50226, AND
TITLE OF INVENTION: 58764,
TITLE OF INVENTION: NOVEL HUMAN TRANSFERASE FAMILY MEMBERS AND USES THEREFOR
FILE REFERENCE: 38155-20034.00
CURRENT APPLICATION NUMBER: US/09/924,358
CURRENT FILING DATE: 2001-08-06
PRIOR APPLICATION NUMBER: US 60/229,300
PRIOR FILING DATE: 2000-09-01
NUMBER OF SEQ ID NOS: 52
SOFTWARE: FastSeq for Windows Version 4.0
SEQ ID NO 29
LENGTH: 150
TYPE: PRT
ORGANISM: Artificial Sequence
FEATURE:
OTHER INFORMATION: Consensus amino acid
US-09-924-358-29

Query Match 12.4%; Score 218; DB 10; Length 150;
Best Local Similarity 40.0%; Pred. No. 2.2e-14;
Matches 52; Conservative 26; Mismatches 44; Indels 8; Gaps 5;

```
QY      6 SVLLCAVNAEYFPGSSAAVVG--TWNRLLDLDGDSSTGGTPRIANPFCOOPDRITIS 64
       1 :::: |::| :::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
Db     1 SIIPPTNEKRYLEECESLNTQTTFENEFELIVDDSGTGTVEILREY -AKDPRIHIVR 59

QY      65 NPRRLGFASINIGLDELANSGGSEYIARTDADDIASPCWTEKVGMENDRSITANGAW 124
       1 :||| ||| |::: |::: |::: |::: |::: |::: |::: |::: |::: |::: |
Db     60 LEENIGLAAAANAALKH--ATGDYDIYAFLDADD-EYVDMLKEELLEERKNGADVIYG-- 114

QY      125 LEVLSENNRK 134
       1:::|
Db     115 -RVINENKGR 123
```

```

RESULT 6
US-09-879-959-10
: Sequence 10. Application US/09879959
: Patent No. US20020160489A1
: GENERAL INFORMATION:
: APPLICANT: Weigel, Paul H
: APPLICANT: Kumari, Kshama
: APPLICANT: DeAngelis, Paul
: TITLE OF INVENTION: STREPTOCOCCUS EQUISIMILIS
: TITLE OF INVENTION: IN BACILLUS SUBTILIS
: FILE REFERENCE: 3554.049
: CURRENT APPLICATION NUMBER: US/09/879,959
: CURRENT FILING DATE: 2001-09-12
: PRIOR APPLICATION NUMBER: 09/469,200
: PRIOR FILING DATE: 1999-12-21
: PRIOR APPLICATION NUMBER: 08/178,851
: PRIOR FILING DATE: 1998-10-26
: NUMBER OF SEQ ID NOS: 10
: SOFTWARE: PatentIn version 3.1
: SEQ ID NO 10
: LENGTH: 972
: TYPE: PRT
: ORGANISM: pasteurella multocida
: US-09-879-959-10

```

Query Match	11.8%	Score 209;	DB 9;	Length 972;
Best Local Similarity	21.8%	Pred. No. 2.5e-12;		
Matches 96;	Conservative 60;	Mismatches 146;	Indels 138;	Gaps 14;

[illegible]

```

RESULT 7
US-09-767-041-22
? Sequence 22, Application US/09767041
? Patent No. US20020055168A1
? GENERAL INFORMATION:
? APPLICANT: Smith, Hilda
? TITLE OF INVENTION: STREPTOCOCCUS SUI S VACCINES AND DIAGNOSTIC TESTS
? FILE REFERENCE: 2183-4776
? CURRENT APPLICATION NUMBER: US/09/767, 041
? CURRENT FILING DATE: 2001-01-22
? PRIOR APPLICATION NUMBER: PCT/NL99/00460
? PRIOR FILING DATE: 1999-07-19
? PRIOR APPLICATION NUMBER: EP98202465.5
? PRIOR FILING DATE: 1998-07-22
? PRIOR APPLICATION NUMBER: EP98202467.1
? PRIOR FILING DATE: 1998-07-22
? NUMBER OF SEQ ID NOS: 53
? SOFTWARE: PatentIn version 3.0
? SEQ ID NO 22
? LENGTH: 332
? TYPE: PR1
? ORGANISM: Streptococcus suis
? FEATURE:
? NAME/KEY: misc_feature
? OTHER INFORMATION: CFS2K
US-09-767-041-22

```

Query Match	11.8%	Score 208;	DB 10;	Length 332;
Best Local Similarity	22.7%	Pred. No. 6.7e-13;		
Matches 82;	Conservative 74;	Mismatches 114;	Indels 92;	Gaps 18

[illegible]

```

RESULT 8
US-09-767-041-36
; Sequence 36, Application US/09767041
; Patent No. US20020055168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; FILE OF INVENTION: STREPTOCOCCUS SOIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4726
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460

```

PRIOR FILING DATE: 1999-07-19
 PRIOR APPLICATION NUMBER: EP96202465.5
 PRIOR FILING DATE: 1998-07-22
 PRIOR APPLICATION NUMBER: EP98202467.1
 PRIOR FILING DATE: 1998-07-22
 NUMBER OF SEQ ID NOS: 53
 SOFTWARE: PatentIn version 3.0
 SEQ ID NO 36
 LENGTH: 278
 TYPE: PRT
 ORGANISM: Streptococcus suis
 FEATURE:
 NAME/KEY: misc-feature
 OTHER INFORMATION: CPSiK
 OS-09-767-041-36

Query Match	11.8%	Score 207.5	DB 10	Length 278
Best Local Similarly	25.28%	Pred. No. 5.8e-13		
Matches 62; Conservative	56;	Mismatches 73;	Indels 55;	Gaps 12

[illegible]

RESULT 9
 US-09-816-028A-31
 Sequence 31, Application US/09816028A
 Patent No. US20020042369A1
 GENERAL INFORMATION:
 APPLICANT: Gilbert, Michel
 APPLICANT: Wakarchuk, Warren W.
 APPLICANT: National Research Council of Canada
 TITLE OF INVENTION: Campylobacter Glycosyltransferases for Biosynthesis of
 Gangliosides and Ganglioside Mimics
 FILE REFERENCE: 019633-0001105
 CURRENT APPLICATION NUMBER: US/09/816,028A
 PRIOR FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: US 60/118,213
 PRIOR FILING DATE: 1999-02-01
 PRIOR APPLICATION NUMBER: US 09/495,406
 PRIOR FILING DATE: 2000-01-31
 NUMBER OF SEQ ID NOS: 49
 SOFTWARE: SeqIdN Ver. 2.1
 SEQ ID NO 31
 LENGTH: 297
 TYPE: PRT
 ORGANISM: Campylobacter jejuni
 FEATURE:
 OTHER INFORMATION: beta-1,3 galactosyl transferase from C. jejuni O:10
 US-09-816-028A-31

	Query Match	11.0%;	Score 194;	DB 10;	Length 297;
	Best Local Similarity	21.8%;	Pred. NC. 1.4e-11;		
	Matches 81;	Conservative 56;	Mismatches 108;	Indels 126;	Gaps 14
07	5	VSVLICANNAERYFPAQSLAAVVGQWIRNIDILIVDGSYDGPALARHFOEDGRIIRITS	64		

QY 5 VSLICAYNAEKYFAQSLAVVGQTRNLDILVDDGSTDGTPAIARHFEQEDGRIRIS 64

[illegible]

```

      RESULT 10
US-09-767-041-35
; Sequence 35, Application US/09767041
; Patent No. US2002005168A1
; GENERAL INFORMATION:
; APPLICANT: Smith, Hilda
; TITLE OF INVENTION: STREPTOCOCCUS SUIIS VACCINES AND DIAGNOSTIC TESTS
; FILE REFERENCE: 2183-4776
; CURRENT APPLICATION NUMBER: US/09/767,041
; CURRENT FILING DATE: 2001-01-22
; PRIOR APPLICATION NUMBER: PCT/NL99/00460
; PRIOR FILING DATE: 1999-07-19
; PRIOR APPLICATION NUMBER: EP9802465.5
; PRIOR FILING DATE: 1998-07-22
; PRIOR APPLICATION NUMBER: EP9802467.1
; PRIOR FILING DATE: 1998-07-22
; NUMBER OF SEQ ID NOS: 33
; SOFTWARE: PatentIn version 3.0
; SEQ ID NO: 35
; LENGTH: 322
; TYPE: PRT
; ORGANISM: Streptococcus suis
; FEATURE:
; NAME/KEY: misc_feature
; OTHER INFORMATION: CPSIJ
US-09-767-041-35

```

Query Match	10.7%;	Score 188;	DB 10;	Length 322;
Best Local Similarity	23.4%;	Pred. No. 6e-11;		
Matches	81;	Conservative	55;	Mismatches 132;
				Indels 78;
				Gaps 15;

[illegible]

Query Match	10.5%;	Score 185.5;	DB 10;	Length 120;
Best Local Similarity	31.2%;	Pred. NO. 2.6e-11;		
Matches 39;	Conservative 29;	Mismatches 38;	Indels 19;	Gaps 2

Db 4 VSLVPIFNTKYLRECLDLSISQSYTNLEILLIDGSSDSDTCLLEAEDGRITKLR 63
 QY 63 -----ISNPNLGFASLNLIGDELAKSGGGEYIARTDADIASPGWIKIYGEKDRS 117
 Db 64 LKNGVSNMRNGI-----KNSNTANIMFVSDSDIDVGNIVESLYTCLKENDS 111
 QY 118 IAMG 122
 Db 112 DLSCG 116

RESULT 14

US-09-816-028A-39
 ; Sequence 39, Application US/09816028A
 ; Patent No. US20020042369A1
 ; GENERAL INFORMATION:
 ; APPLICANT: Gilbert, Michel
 ; APPLICANT: Wakarchuk, Warren W.
 ; APPLICANT: National Research Council of Canada
 ; TITLE OF INVENTION: Campylobacter glycosyltransferases for Biosynthesis of
 ; FILE REFERENCE: 019633-000111US
 ; CURRENT APPLICATION NUMBER: US/09/816,028A
 ; CURRENT FILING DATE: 2001-03-21
 ; PRIOR APPLICATION NUMBER: US 60/118,213
 ; PRIOR FILING DATE: 1999-02-01
 ; PRIOR APPLICATION NUMBER: US 09/495,406
 ; PRIOR FILING DATE: 2000-01-31
 ; NUMBER OF SEQ ID NOS: 49
 ; SOFTWARE: Patent In Ver. 2.1
 ; SEQ ID NO 39
 ; LENGTH: 270
 ; TYPE: PRP
 ; ORGANISM: Campylobacter jejuni
 ; FEATURE:
 ; OTHER INFORMATION: glycosyltransferase from C. jejuni OH4384 (ORF 12a)
 ; OTHER INFORMATION: of lipooligosaccharide (LOS) biosynthesis locus)
 US-09-816-028A-39

Query Match 10.5%; Score 185.5; DB 10; Length 270;
 Best Local Similarity 21.3%; Pred. No. 8,3e-11;
 Matches 65; Conservative 48; Mismatches 79; Indels 113; Gaps 10;

QY 3 PLVSVLCAVNAEKYPAQSLAAVVGQTMRLDILLVDDGSDTGTATIAHFOEDGRIRI 62
 Db 2 PQSLIILPLFNSCDFISRLQSCINOTLMDIELLIDDKSKNSLWVLEFAKKDRIRI 61
 QY 63 ISNPNLGFASLNLIGDELAKSGGGEYIARTDADIASPGWIKIYGEKDRSIIAMG 122
 Db 62 FQNEMLGTFASRLGV-----LHSSDFIMFLSDDFLTPDACEIAFKEMKKGFDLLC-- 115
 QY 123 AMLEVSEENNSVLAIAIRNGAIWPKTRHEDIVAVFPFGNPIHNTMIRSVVDGL 182
 Db 116 ----- 115
 QY 183 REDPAIHAEDYKFWYEAQGLRLAYVPEALVYRPHODOTS-----KYNLQORRTAMK 237
 Db 116 -ED-ATVHVHVKTKF-----YRKQDEVFQKQLEFLSLQNRHFCMS 155
 QY 238 IKEEIRAGIWKAGIAGVADCLWYG-LKSTAYALYERKALS-GQDIGICRLFLYEFLSL 295
 Db 156 V-----W-----AKCFKDIILIKSEFKIKIDERLNGEDV-----LFCYIYFMFC 195
 QY 296 EKYSL 300
 Db 196 EKIAV 200

RESULT 15
 US-09-765-272-196
 ; Sequence 196, Application US/09765272
 ; Patent No. US20020061545A1
 ; GENERAL INFORMATION:

APPLICANT: Choi et. al.
 ; TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
 ; NUMBER OF SEQUENCES: 452
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: Human Genome Sciences, Inc.
 ; STREET: 9410 Key West Avenue
 ; CITY: Rockville
 ; STATE: Maryland
 ; COUNTRY: USA
 ; ZIP: 20850
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
 ; COMPUTER: HP Vectra 486/33
 ; OPERATING SYSTEM: MSDOS version 6.2
 ; SOFTWARE: ASCII Text
 ; CURRENT APPLICATION DATA:
 ; APPLICATION NUMBER: US/09/765,272
 ; FILING DATE: 22-Jan-2001
 ; CLASSIFICATION: <Unknown>
 ; PRIOR APPLICATION DATA:
 ; APPLICATION NUMBER: 08/961,083
 ; FILING DATE: <Unknown>
 ; ATTORNEY/AGENT INFORMATION:
 ; NAME: Brookes, A. Anders
 ; REGISTRATION NUMBER: 36,373
 ; REFERENCE/DOCKET NUMBER: PB340P2
 ; TELECOMMUNICATION INFORMATION:
 ; TELEPHONE: (301) 309-8504
 ; TELEFAX: (301) 309-8512
 ; INFORMATION FOR SEQ ID NO: 196:
 ; SEQUENCE CHARACTERISTICS:
 ; LENGTH: 281 amino acids
 ; TYPE: amino acid
 ; STRANDEDNESS: single
 ; TOPOLOGY: linear
 ; MOLECULE TYPE: protein
 ; SEQUENCE DESCRIPTION: SEQ ID NO: 196:
 US-09-765-272-196

Query Match 10.5%; Score 185.5; DB 10; Length 281;
 Best Local Similarity 23.1%; Pred. No. 8,7e-11;
 Matches 75; Conservative 53; Mismatches 118; Indels 79; Gaps 13;

QY 25 VVGQTRNLDILLVDDGSDTGTATIAHFOEDGRIRISNPNLGFASLNLIGDELAK 84
 Db 5 ILKQYQNIETIILVDDGSDTNGEIGDAPMQRKRVVLRHGNKGAQAQAKMGI-SYAK 63
 QY 85 SGGGEYIARTDADIASPGWIKIYGEKDRSIIAMGAMLEVSEENN----- 133
 Db 64 ---GEYITVSDSDIVKENMIEFLVQVQEKDADVIGYVYVDESDGNFYFYVGQDFC 120
 QY 134 ---KSVLAIAIRNGAIWPKTRHEDIVAVF-----PFGNPIHNTMIRSVI DGLRD 185
 Db 121 VEEALQEIIMNRQAGW-KFNSSAFILPFEKLIKELFNEVH-----FSNGRRRD 169
 QY 186 PAYIHAEDYKFWYEAQGLRLAYVPEALVYRPHODOTSRYNLQORRTAMKIKEIRAG 245
 Db 170 D---EATMHRFILLASK---IYFINDNLVLR-----RRGSIIMRREFDLS 209
 QY 246 YKKAAGIAGV---ADCLVYGLKSTAYALYERKALS-GQDIGICRL---FLYEFLSLEK 297
 Db 210 -WARDIVEYFSKISDCV-----LAGLDVSVLRIRFVNLADVYQOTLEY 252
 QY 298 YSLTDLDFLTDRVARKLFAPOYR 322
 Db 253 HOLITREYKDCICFRLLKLFDAEOR 277

Search completed: December 2, 2002, 12:16:02
 Job time : 7.75447 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 14.9563 seconds
(Without alignments)
2166.126 Million cell updates/sec

Title: US-10-007-267-5
Perfect score: 1764
Sequence: 1 MQLPLSVLICAYNAEKYFAQ.....APQYRKILKKMLPWWKYSY 337

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: PIR_73: *
2: PIR1: *
3: PIR2: *
4: PIR3: *

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1074.5	60.9	333	1 S70813	glycosyl transferase
2	1069.5	60.6	349	2 D81027	lacto-N-neotetraos
3	1039	58.9	346	2 H81970	lacto-N-neotetraos
4	682.5	38.7	323	1 H64130	glycosyl transferase
5	346	19.6	367	2 G95948	probable glycosylt
6	282.5	16.0	333	2 B97168	glycosyltransferase
7	263.5	14.9	336	2 A97168	glycosyltransferase
8	258.5	14.7	340	2 T44330	glycosyl transferase
9	258	14.6	333	2 H97167	glycosyltransferase
10	249	14.1	271	2 B84114	exopolysaccharide
11	248.5	14.1	290	2 H64431	glycosyl transferase
12	244.5	13.9	343	2 A12091	glucosyltransferase
13	243	13.8	298	2 B75096	glycosyl transferase
14	234	13.3	298	2 T44330	glycosyl transferase
15	231	13.1	278	2 AG1920	glycosyl transferase
16	226.5	12.8	301	2 D70036	exopolysaccharide
17	225.5	12.8	301	2 F95205	glycosyl transferase
18	225	12.8	318	2 AG2189	hypothetical prote
19	225	12.8	334	1 G71153	hypothetical prote
20	224	12.7	732	2 A84107	hypothetical prote
21	223.5	12.7	311	2 T00087	thiamosyltransferase
22	223	12.6	321	2 AG2188	hypothetical prote
23	223	12.6	333	2 AH2036	hypothetical prote
24	221	12.5	344	2 G70036	spore coat polysac
25	220	12.5	315	2 T44648	glycosyl transferase
26	219.5	12.4	294	2 E83022	probable glycosyl
27	219.5	12.4	324	2 A69290	glycosyl transferase
28	218.5	12.4	322	2 T44647	hypothetical prote
29	216.5	12.3	323	2 AD2189	hypothetical prote

30	212.5	12.0	344	2 AC0974	probable glycosylt
31	212	12.0	337	2 AD3614	glycosyl transferase
32	211.5	12.0	257	2 E84107	teichuronic acid b
33	209	11.8	250	2 A64059	glycosyl transferase
34	209	11.8	972	2 T09595	glucuronosyltransf
35	207.5	11.8	281	2 G97777	glycosyl transferase
36	207	11.7	604	2 E97757	hypothetical prote
37	206.5	11.7	318	2 T50039	beta-1,4-galactosy
38	205.5	11.6	318	1 E71690	minor teichoic aci
39	205	11.6	328	2 F98024	raffinose-raffinose
40	204.5	11.6	324	2 AB2190	hypothetical prote
41	203.5	11.5	280	2 E71703	glycosyl transferase
42	203.5	11.5	327	2 AB1211	glycosyltransferase
43	202	11.5	330	2 AH2188	hypothetical prote
44	201	11.4	299	2 B83557	probable glycosyl
45	199	11.3	328	2 F95158	glycosyl transferase

ALIGNMENTS

RESULT 1
S70813
glycosyl transferase A (EC 2.4.-.-) - Neisseria meningitidis
C:Species: Neisseria meningitidis
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 21-Jul-2000
C:Accession: S70813
R:Jennings, M.P.; Hood, D.W.; Peak, I.R.A.; Virji, M.; Moxon, E.R.
Mol. Microbiol. 18, 729-740, 1995
A:Title: Molecular analysis of a locus for the biosynthesis and phase-variable expres
A:Reference number: S70812; MUID:96414473; PMID:8817494
A:Accession: S70813
A:Status: preliminary; nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-333 <JEN>
A:Cross-references: EMBL:U25839; NID:9973183; PIDN:AAC44084.1; PID:9973185
A:Note: the nucleotide sequence was submitted to the EMBL Data Library, April 1995
C:Genetics:
A:Gene: lgtA
A:Start codon: TTG
C:Superfamily: Neisseria meningitidis glycosyl transferase A
C:Keywords: glycosyltransferase

Query Match	60.9%;	Score 1074.5;	DB 1;	Length 333;
Best Local Similarity	65.9%;	Pred. No. 1.2e-81;		
Matches 218;	Conservative 29;	Mismatches 83;	Indels 1;	Gaps 1
QY	1	MQLPLSVLICAYNAEKYFAQSLAAVAGOTWNLDTLIVDGSSTGTPAIRHFOFODGRI	60	
Db	1	MQLPLSVLICAYNAEKYFAQSLAAVAGOTWNLDTLIVDGSSTGTPAIRHFOFODGRI	60	
QY	61	RIISPNRLGFIASINIGLDELAKS-GGGEYIARTDADDIASPGWIEKIVGEMEKRSII	119	
Db	61	RIISPNRLGFIASINIGLDELAKS-GGGEYIARTDADDIASPGWIEKIVGEMEKRSII	120	
QY	120	AMGAMLEVLSEENKSVLAATARNCAINDKPTRHEDIYAVPPFGPHINNTMKRSYID	179	
Db	121	AMGAMLEVLSEENKSVLAATARNCAINDKPTRHEDIYAVPPFGPHINNTMKRSYID	180	
QY	180	GGLEFDPAYIAEDYKFEYEAAGKGLRLAYPEALVKKYFHDQTSKKNLOORTAMRK	239	
Db	181	GGLEFDPAYIAEDYKFEYEAAGKGLRLAYPEALVKKYFHDQTSKKNLOORTAMRK	240	
QY	240	EIRAGYKKAAGIAGVADCLNYGLKSTAYALYKALSGODIGCLRLVYFLSLEKYS	299	
Db	241	ETARNDLFQSGKGFTRPDSLEYRQTKAAVAYELLEKHLPEDELARRLFYOCFKRTDLP	300	
QY	300	LTDLDLFTDVMRKRLFAAPQYRKILKKMLR	330	
Db	301	AGAMLDFAADGMRRLFTLRLQYFGILHRLK	331	
RESULT	2			

D81027
lacto-N-neotetraose biosynthesis glycosyl transferase Lgta NMB1929 [imported] - Neisseria
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: DB1027
R:Retelien, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen, J.A.
Hickley, E.K.; Hall, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.A.;
Hill, H.; Qin, H.; Vamathevan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Plaza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappuoli, R.; Ve
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: AB1000; MUID:20175755; PMID:10710307
A:Accession: DB1027
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-349 <TEXT>
A:Cross-references: GB:AE002541; GB:AE002098; NID:g7227175; PIDN:AAF42528.1; PID:g722718
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1929
C:Superfamily: Neisseria meningitidis glycosyl transferase A
Query Match 60.6%; Score 1069.5; DB 2; Length 349;
Best Local Similarity 65.6%; Pred. No. 3.2e-81;
Matches 217; Conservative 30; Mismatches 83; Indels 1; Gaps 1;
QY 1 MOPLYSVLICAVNAEKYFAQSIAAIVGOTWRNLDILVDDGSTGTPAIRHFOEDGR 60
DB 17 LQPLVSVLICAVNAEKYFAQSIAAIVGOTWRNLDILVDDGSTGTPAIRHFOEDGR 76
QY 61 RIISPRNLGFASINIGLDELAKS-GGGEYIARTDADDIASPGWIEKIVGEMEKDRSTI 119
DB 77 RLADPRNSGLPSINIGLDELAKSGGGEYIARTDADDIAAPWIEKIVGEMEKDRSTI 136
QY 120 AMGAWLEVLSEENKSVLAALARNGAIMDKPTRHEDIYAVFPFGNPINHTMIRRSYD 179
DB 137 AMGAWLEVLSEENKSVLAALARNGAIMDKPTRHEDIYAVFPFGNPINHTMIRRSYD 196
QY 180 GGLRFPAYIAHEDYKFWYKAGLGRALVYPRALVYKPRPHODQTSKYNLQOORTAWKIK 239
DB 197 GGLRFPAYIAHEDYKFWYKAGLGRALVYPRALVYKPRPHODQTSKYNLQOORTAWKIK 256
QY 240 BEIRAGYKKAAGINAGACINLGLKSTAYALYKALSGODIGCLRFLYEFYLSLEKYS 299
DB 257 KTARNDFLOSMGFKTRFDSLEYRQIKAVAYELLEKHLPEDERARFLYOCFKRTDPLP 316
QY 300 LTDLDELDRVWRKLFAPAYRKILKMLR 330
DB 317 AGAWLDFPADGMRRLFTLRQYFGLHRLK 347
RESULT 3
H81970
lacto-N-neotetraose biosynthesis glycosyl transferase NMA0524 [imported] - Neisseria men
C:Species: Neisseria meningitidis
C:Date: 05-May-2000 #sequence_revision 05-May-2000 #text_change 02-Feb-2001
C:Accession: H81970
R:Parkhill, J.; Achtman, M.; James, K.D.; Bentley, S.D.; Churcher, C.; Klee, S.R.; Morel
; Holtroyd, S.; Jørgensen, K.; Leather, S.; Moulé, S.; Mungall, K.; Quail, M.A.; Rajadream,
Nature 404, 502-506, 2000
A:Title: Complete DNA sequence of a serogroup A strain of Neisseria meningitidis 22491.
A:Reference number: AB1775; MUID:20222556; PMID:10761919
A:Accession: H81970
A:Status: Preliminary
A:Molecule type: DNA
A:Residues: 1-346 <PAR>
A:Cross-references: GB:AL162753; GB:AL157959; NID:g7379120; PIDN:CAB83816.1; PID:g737926
A:Experimental source: serogroup A, strain 22491
C:Genetics:
A:Gene: Lgta; NMA0524
C:Superfamily: Neisseria meningitidis glycosyl transferase A
Query Match 58.9%; Score 1039; DB 2; Length 346;

Best Local Similarity 63.3%; Pred. No. 1.1e-78;
Matches 209; Conservative 35; Mismatches 86; Indels 0; Gaps 0;
QY 1 MOPLYSVLICAVNAEKYFAQSIAAIVGOTWRNLDILVDDGSTGTPAIRHFOEDGR 60
DB 1 MOPLYSVLICAVNAEKYFAQSIAAIVGOTWRNLDILVDDGSTGTPAIRHFOEDGR 60
QY 61 RIISPRNLGFASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSTI 120
DB 61 RIISPRNLGFASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSTI 120
QY 121 MGAWLEVLSEENKSVLAALARNGAIMDKPTRHEDIYAVFPFGNPINHTMIRRSYD 180
DB 121 MGAWLEVLSEENKSVLAALARNGAIMDKPTRHEDIYAVFPFGNPINHTMIRRSYD 180
QY 181 GGLRFPAYIAHEDYKFWYKAGLGRALVYPRALVYKPRPHODQTSKYNLQOORTAWKIK 240
DB 181 GGLRFPAYIAHEDYKFWYKAGLGRALVYPRALVYKPRPHODQTSKYNLQOORTAWKIK 240
QY 241 EIRAGYKKAAGINAGACINLGLKSTAYALYKALSGODIGCLRFLYEFYLSLEKYS 300
DB 241 EIRAGYKKAAGINAGACINLGLKSTAYALYKALSGODIGCLRFLYEFYLSLEKYS 300
QY 301 LTDLDELDRVWRKLFAPAYRKILKMLR 330
DB 301 LTDLDELDRVWRKLFAPAYRKILKMLR 330
RESULT 4
H64130
glycosyl transferase homolog H11578 - Haemophilus influenzae (strain Rd KW20)
C:Species: Haemophilus influenzae
C:Date: 10-Sep-1999 #sequence_revision 10-Sep-1999 #text_change 28-Jul-2000
C:Accession: H64130
R:Fleischmann, R.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Vente
A:Title: Whole-genome random sequencing and assembly of Haemophilus influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: H64130
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-323 <TTGR>
A:Cross-references: GB:U02832; GB:L42023; NID:g1574421; PIDN:AAC23227.1; PID:g1574422
C:Superfamily: Neisseria meningitidis glycosyl transferase A
Query Match 38.7%; Score 682.5; DB 1; Length 323;
Best Local Similarity 45.9%; Pred. No. 4.3e-49;
Matches 151; Conservative 54; Mismatches 107; Indels 17; Gaps 6;
QY 3 PLYSVLICAVNAEKYFAQSIAAIVGOTWRNLDILVDDGSTGTPAIRHFOEDGR 62
DB 5 PLYSVLICAVNAEKYFAQSIAAIVGOTWRNLDILVDDGSTGTPAIRHFOEDGR 64
QY 63 ISNPNLGFASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSTI 122
DB 63 ISNPNLGFASINIGLDELAKSGGGEYIARTDADDIASPGWIEKIVGEMEKDRSTI 122
QY 123 AMLEVLSEENKSVLAALARNGAIMDKPTRHEDIYAVFPFGNPINHTMIRRSYD 181
DB 123 AMLEVLSEENKSVLAALARNGAIMDKPTRHEDIYAVFPFGNPINHTMIRRSYD 181
QY 181 GGLRFPAYIAHEDYKFWYKAGLGRALVYPRALVYKPRPHODQTSKYNLQOORTAWKIK 241
DB 181 GGLRFPAYIAHEDYKFWYKAGLGRALVYPRALVYKPRPHODQTSKYNLQOORTAWKIK 241
QY 241 EIRAGYKKAAGINAGACINLGLKSTAYALYKALSGODIGCLRFLYEFYLSLEKYS 301
DB 241 EIRAGYKKAAGINAGACINLGLKSTAYALYKALSGODIGCLRFLYEFYLSLEKYS 301
QY 302 LTDLDELDRVWRKLFAPAYRKILKMLR 330

Db 291 SLHPI--KYLHEFDLKNKIKRKR 317

RESULT 5

G95948

Probable glycosyltransferase protein SMB21189 [Imported] - Sinorhizobium meliloti (strain C:Species: Sinorhizobium meliloti)

C:Date: 24-Aug-2001 #sequence_revision 24-Aug-2001 #text_change 30-Sep-2001

C:Accession: G95948

R:Finan, T.M.; Weidner, S.; Wong, K.; Buhmester, J.; Chain, P.; Vorholter, F.J.; Hernat Proc. Natl. Acad. Sci. U.S.A. 98, 9889-9894, 2001

A:Title: The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-fixing endo A:Reference number: A95842; MUID:21396508; PMID:11481431

A:Accession: G95948

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-367 <KUR>

A:Cross-references: GB:AL591985; PIDN:CAC49255.1; PID:915140741; GSPDB:GN00167

A:Experimental source: strain 1021, megaplasmid pSymb

R:Galibert, F.; Finan, T.M.; Long, S.R.; Puhler, A.; Abola, P.; Ampe, F.; Barloy-Hubler, P.; Chain, P.; Cowie, A.; Davis, R.W.; Dreano, S.; Federspiel, N.A.; Fisher, R.F.; L.; Hyman, R.W.; Jones, T.

Science 293, 668-672, 2001

A:Authors: Kahn, D.; Kahn, M.L.; Kalman, S.; Keating, D.H.; Kiss, E.; Komp, C.; Lelaure, hebaull, P.; Vandenbol, M.; Vorholter, F.J.; Weidner, S.; Wells, D.H.; Wong, K.; Yeh, K.

A:Title: The composite genome of the legume symbiont Sinorhizobium meliloti.

A:Reference number: A96039; MUID:21368234; PMID:11474104

A:Contents: annotation

C:Genetics:

A:Gene: SMB21189

A:Genome: plasmid

Query Match 19.6%; Score 346; DB 2; Length 367;
Best Local Similarity 30.7%; Pred. No. 4.9e-21;

Matches 112; Conservative 53; Mismatches 130; Indels 70; Gaps 13;

3 PLVSLICAYNAEKYFAQSLAAVVGOTWRNLDILIVDGSSTGTPAIAAHFQODGRIRI 62

4 PVYSIVLPYNAEPYIAAIESVLQDYERLEIVAIMDGSSTDRSDILERYKKSRSVSI 63

63 ISNPNLGFASINIGDELAKSGGGEYIARTDADIASPGWIEKIVGEMKRSIIAMG 122

64 ISR-ENRGIATINLNGEL-ALAK---GELIARMDADDIAPSRSLSQVALFSAEPRLALSG 118

123 AMLEVLSSENNKSVLAALARNCAIMDKPTRHEDIYAVFPFGNPH----- 167

119 TGIDML-----IGNRIIRGKP-----NPIYRPSGLRILISMEFTIF 153

168 -NNTIMIRRSVI-DGGLRFPAYIAHEDYKFEWYEAQGLRLAYPEALVYKRFHODQTS 225

154 MHSIVYRNRYVPEMLRNDPVAHEDFDLFRITADRPVIMIDEALVAVYIHDSVTS 213

226 KYNLQORTAMKIKEIIRA-----GYKAAAGIAGVADCLNGLSTAYALYEKALS 277

214 KHKRMRTHLIVAEINARDAIRLDSALAEIGAIVTSFV--ARLADLVIAL-EREIS 270

278 GO-----DIG--CLRLFEYFLSLEKYSITDLDLDFIDRWARKLEAFOYRKILKK 327

271 AOPGEVRRAYEDGALCFYFLYQLIAEBEOPRLTH--EFLITRTGKWGLIRRRERYGLAA 328

328 MLRPW 332

329 ARAPW 333

RESULT 6

B97168

glycosyltransferase [Imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: B97168

R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; Lee,

; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: B97168

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-333 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80133.1; PID:915025169; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2175

Query Match 16.0%; Score 282.5; DB 2; Length 333;
Best Local Similarity 27.4%; Pred. No. 8.1e-16;

Matches 80; Conservative 59; Mismatches 102; Indels 51; Gaps 10;

4 LVSVLICAYNAEKYFAQSLAAVVGOTWRNLDILIVDGSSTGTPAIAAHFQODGRIRI 63

1 MISVIMPYNCERKYLEESIESILKOTYRDFEFIIYVDSGSDKSIDINKYANDNDRIYVY 60

64 SNPNLGFASINIGDELAKSGGGEYIARTDADIASPGWIEKIVGEMKRSIIAMKA 123

61 SDDNNKGMVYSLNEGIDR-AK---GSYVARMADADIALPERFERQIEYLKMKKDVIIAC 116

124 WLEVLSSENNKSVLAALARNCAIMDKPTRHEDIYAVFPFGNPIHNTMIRRSVI 178

117 KYEAGDVSREKLEREMHYNDLNS-----ESISLELNCYIAHPSVWKKMSVL 168

179 D--GGLRFPAYIAHEDYKFEWYEAQGL-RLAYPEALVYKRFHODQ-----SSKY 227

169 KALGGYNLNL--YKRTEDYVNLMLRAIAKGYKIAMLEKIMKIRLHNDKTHRDAEGFSIR 226

228 NLOQORTAMKIKEIIRAGYKWAAGIAGVADCLNGLSTAYALYEKALSQ 279

227 DIIOQRLEY-VREKLR-----LQDSYIVGASNGK 257

RESULT 7

A97168

glycosyltransferase [Imported] - Clostridium acetobutylicum

C:Species: Clostridium acetobutylicum

C:Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001

C:Accession: A97168

R:Noelling, J.; Breton, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Gibson, R.; L

; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.

J. Bacteriol. 183, 4823-4838, 2001

A:Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium

A:Reference number: A96900; MUID:21359325; PMID:21359325

A:Accession: A97168

A:Status: preliminary

A:Molecule type: DNA

A:Residues: 1-336 <KUR>

A:Cross-references: GB:AE001437; PIDN:AAK80132.1; PID:915025168; GSPDB:GN00168

A:Experimental source: Clostridium acetobutylicum ATCC824

C:Genetics:

A:Gene: CAC2174

Query Match 14.9%; Score 263.5; DB 2; Length 336;
Best Local Similarity 28.2%; Pred. No. 3.1e-14;

Matches 87; Conservative 61; Mismatches 123; Indels 37; Gaps 13;

3 PLVSLICAYNAEKYFAQSLAAVVGOTWRNLDILIVDGSSTGTPAIAAHFQODGRIRI 62

5 PVYSIVLPYNAEPYIAAIESVLQDYERLEIVAIMDGSSTDRSDILERYKKSRSVSI 64

63 ISNPNLGFASINIGDELAKSGGGEYIARTDADIASPGWIEKIVGEMKRSIIAMG 122

65 ISR-EHRGIVDSLNGIN-IAR---GKIYARMDADDISINNIEKQEFLELKNQVDILG 119

123 AMLEVLS--EENNKSVLAALARNCAIMDKPTRHEDIYAVFPFGNPI-HNNTMIRRSYID 179

120 TRIEAFGIDIDEKOKITY-----NSAFSIR-FDSQNIQVFLTSCAIIHPSVWFKKDSIVK 173

[illegible]

RESULT 8
T44330
glycosyl transferase homolog [imported] - *Vibrio cholerae*
C:Species: *Vibrio cholerae*
C:Date: 21-Jan-2000 #sequence_revision 21-Jan-2000 #text_change 21-Jul-2000
C:Accession: T44330
R:Ryamasaki, S.; Shimizu, T.; Hoshino, K.; Ho, S.T.; Shimada, T.; Nair, G.B.; Takeda, Y.
Gene 237, 321-332, 1999
A:Title: The genes responsible for O-antigen synthesis of *Vibrio cholerae* O139 are close
A:Reference number: Z22749; MUID:99453293; PMID:10521656
A:Accession: T44330
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: DNA
A:Residue: 1-340 <YAM>
A:Cross-references: EMBL:AB012957; NID:g4115688; PIDD:BAA33634.1; PTD:g3721684
A:Experimental source: strain 022
C:Genetics:
A:Note: wblc
C:Superfamily: *Neisseria meningitidis* glycosyl transferase A

Query Match	14.7%	Score 258.5	DB 2	Length 340
Best Local Similarity	30.0%	Pred. No. 8.3e-14		
Matches 69	Conservative 45	Mismatches 95	Indels 21	Gaps 5

QY	3	P L V S Y L I C A Y A A E K Y F A O S L A A V G C T W R N I D I L I V D D G S T D G T F A I A R H E O D G R I R I	62
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	6	P T I S Y I M S V N G E K I L A Q A I E S I L N Q T S D F E F I I V D G S D T S L I I Q A Y M D K R D R I V L	65

QY	63	I S N F R N D G F I A S I N G L D E L A K S G G G E Y I A R D A D D I A S P G N I E K Y I V E M E K D R S I I M G	122
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	66	I S R V - N K L P Y S L N - - - E A I S V S K A N Y I A R D A D A D I S L P E K L F Q L A V M E N N P I G V G	120

QY	123	A M L E V I S E E N K S V L A A I A R N G A I W D K P T R H E D I Y A A V P E F G N P I H N N T M I M R S Y I D G L	182
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	121	T L A Y L F R E T P S K N K M C - - - - H P E D H D S L I R L F L F S V C F I H P V A Y M I R K S V L D - - -	168

QY	183	R F D P P A Y - - - I H A E D Y K F W Y E N G K I G R L A Y P P E A L V K Y R F H O D O T S S K Y N	228
		: : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :	
Db	169	O L D Y V A N E N F R N S O D Y E L M S R I A E K T R E V Y T I O K P L I F F R O D P D G T S K Y N	218

```

RESULT 9
H97167
glycosyltransferase [imported] - Clostridium acetobutylicum
C.Species: Clostridium acetobutylicum
C.Date: 14-Sep-2001 #sequence_revision 14-Sep-2001 #text_change 14-Sep-2001
C.Accession: H97167
R.Nolling, J.; Breston, G.; Omelchenko, M.V.; Markarova, K.S.; Zeng, Q.; Glibson, R.; Lee,
J.; Daly, M.J.; Bennett, G.N.; Koonin, E.V.; Smith, D.R.
J. Bacteriol. 183, 4823-4838, 2001
A.Title: Genome Sequence and Comparative Analysis of the Solvent-Producing Bacterium Clostridium
A.Reference number: A96900; MUID:21359325; PMID:21359325
A.Accession: H97167
A.Status: preliminary
A.Molecule type: DNA
A.Cross-refs: 1-33 <KUD>
A.Cross-references: GB:AE001437; PIDN:AAK80131.1; PID:G15025167; GSPDB:GN00168
A.Experimental source: Clostridium acetobutylicum ATCC824
A.Genetics:

```

A:Gene: CAC2173
 Query Match 14.6%; Score 258; DB 2; Length 333;
 Best Local Similarity 27.5%; Pred. No. 8.9e-14;
 Matches 95; Conservative 68; Mismatches 123; Indels 60; Gaps 16;

```

Db      4 VSVWVPVYNSEKYLEKESIESILINOSYSPLEFIITINDGSTDSFKIKEYAALDKRINWIS 63
QY      65 NERNLGFASLITNGIDELAKSGGGEYIARTDADDIASPGWIEKTYGEMEKDRSIIAMGAW 124
Db      64 R-ENKGIYLSLWEAL-RLAK---GEYIARMADADDISARKIEKQISFLSKHRDIDICTQ 118
QY      125 LEVLS-----EENKSVLAALARMGAIWDKPTRHIEDIVAAVPPGPNPLPHNNTMIRRSVI 178
Db      119 VKVWNISINDIKERKENKMLNI---EFDIYD-DNREKILNYYCYCLAH---SYMFRKDI- 169
QY      179 DGLGRFPDAY--IHAEDEKFPWEAGKLGRLAY-YPEALVYKRPFHODQS--SKYNLOQR 233
Db      170 ---LHELKGYNDFKESDLDMLIRATESGFKYKKEELIYRHMESSTRVDNONYEGLK 226
QY      234 TAWKI-----KEEIRAGYWKAA-GIANGADCLNYGLKSTAYVALYERAKSGO--- 279
Db      227 DQIKIKLIDVFKRREKDPKPYIWMGASNGKITREVLDEPFERKSQCIAYDVKFTGEPEK 286
QY      280 -----DIGCLRLLY-----EYFSLSEKYSLTDLDPFT 308
Db      287 IRLHPKDIINHRIK-EDYVFIATPEKGEAMSLSKSMGKCIKIDFIS 331

```

RESULT 10
 BH4114
 exopolysaccharide biosynthesis BH3714 [imported] - *Bacillus halodurans* (strain C-125)
 C:Species: *Bacillus halodurans*
 C:Date: 01-Dec-2000 #sequence_revision 01-Dec-2000 #text_change 15-Jun-2001
 C:Accession: BH4114
 R:Takami, H.; Nakasone, K.; Takaki, Y.; Maeno, G.; Sasaki, R.; Masui, N.; Fujii, F.; H
 Nucleic Acids Res. 28, 4317-4331, 2000
 A>Title: Complete genome sequence of the alkaliphilic bacterium *Bacillus halodurans* a
 A:Reference number: A83650; MUID:20512582; PMID:11058132
 A:Accession: BH4114
 A>Status: preliminary
 A:Molecule type: DNA
 A:Residues: 1*271 <STO>
 A:Cross-References: GB:A001519; GB:BA000004; NID:g10176109; PIDN:BAB07433.1; GSPDB:G-G
 A:Experimental source: Strain C-125
 C:Genetics:
 A:Gene: BH3714

	Query Match	14.1%;	Score 249;	DB 2;	length 271;
	Best Local Similarity	26.7%;	Pred. No. 3.8e-13;		
	Matches	70;	Conservative	58;	Mismatches 108; Indels 26; Gaps 9;
QY	1 MNPVSVLICAVNAEKYFAQSLSAAVVGOTWRNLDLLIVDGSSTGCPAIAHFEQDGRI	60			
	: : : : : : : : : : : : : : : : : :				
Db	1 NMPEVTALSVYNDKNYSSESISILNTFENFELLINDSTOGSCGLLEBYSKDKRI	60			
QY	61 RIISVPRLNGFTASINIGDELAKSGGEYIARTADDIASPQWKEIVGEMEDRSIIA	120			
	: : : : : : : : : : : : : : : : : :				
Db	61 RLHHKKNRBGSLYSTLAEGV-SLAKA---PWLRMDADVSRFDRLAYOMDHKAHSELDI	116			
QY	121 MGAWLEVLSEENNKSVLAALAENGAIWDKPRTHEIDIVAF--PFGNDIHNTMIIMRRSV	177			
	: :				
Db	117 LGSYVIDDDIKKENLEIRKV-----PTHKRIANLIWTCP---IHPVLFKKDSI	164			
QY	178 IDCGLRFPAPYATHADGFYWEYA--GKLGRLAYVEALVXYRPHOD---QTSSKYNLQOR	232			
	- - - : : : : : : : : : : : : : : :				
Db	165 IKAQ-SYDRNNLRRODYDLMEFRCLPAKL-KEENDIKPLITYRSTDYDKYKKNNFKVQVOQA	222			
QY	233 RTAMRKIEEIRAGYWKKAAGIAV	254			
	: : : : : : :				
Db	223 KMGFAGARRVRKAPYATIGITV	244			

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 8.20186 Seconds
(without alignments)
1704.189 Million cell updates/sec

Title: US-10-007-267-5

Perfect score: 1764
Sequence: 1 MQLVSVLICAYNAEKYFAQ.....APQYRKILKMLRPMKYRSY 337

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 segs, 41476328 residues

Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0

Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SwissProt_40.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the total score being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	683.5	38.7	323	YF78_HAEIN	Q57287 haemophilus
2	253.5	14.4	290	YAS7_METTA	Q58457 methanococ
3	209	11.8	230	Y868_HAEIN	Q57022 haemophilus
4	195.5	11.1	322	Y586_ANASP	P22639 anabaena sp
5	194	11.0	266	AMSE_ERWAM	Q46635 erwina amy
6	180	10.2	301	AMSB_ERWAM	Q46632 erwina amy
7	178.5	10.1	344	Y1BD_ECOLI	P11290 escherichia
8	167	9.5	446	GGAA_BACSU	P46917 bacillus su
9	164.5	9.3	330	EXOA_RHIME	P33691 rhizobium m
10	160.5	9.1	256	SPSA_BACSU	P33621 bacillus su
11	155	8.8	348	EXOO_RHIME	P33697 rhizobium m
12	150.5	8.5	441	YCDQ_ECOLI	P75905 escherichia
13	147	8.3	268	YMDF_BACSU	P36614 bacillus su
14	142	8.0	346	YF20_MYCTU	Q30587 mycobacteri
15	140.5	7.9	267	YG95_HAEIN	Q46215 haemophilus
16	138.5	7.6	299	Y025_MYCPN	P75086 mycoplasma
17	134.5	7.4	342	GGAB_BACSU	P46918 bacillus su
18	131	7.3	309	EXOU_RHIME	P33700 rhizobium m
19	129.5	7.2	260	YAGI_RHISN	P55465 rhizobium s
20	127	7.1	260	DPML_HUMAN	O70152 homo sapien
21	125	7.0	266	DPML_MOUSE	O70152 mus musculu
22	124	6.9	299	Y060_MYCPN	P75042 mycoplasma
23	122.5	6.7	319	EXOW_RHIME	P33702 rhizobium m
24	118	6.7	424	NODC_RHITO	P17862 rhizobium l
25	117.5	6.7	513	YTH1_RHOER	P46370 rhodococcus
26	117.5	6.4	1275	YFBC_MYXAA	Q30864 myxococcus
27	113.5	6.4	323	YKCC_BACSU	Q34319 bacillus su
28	112.5	6.3	294	YG96_HAEIN	Q48214 haemophilus
29	112	6.3	279	WCAA_ECOLI	P77414 escherichia
30	109.5	6.2	236	DPML_SCHPO	O14466 schizosacch
31	109	6.2	319	YF18_MYCTU	Q30590 mycobacteri
32	109	6.2	413	NODC_RHISN	P50357 rhizobium s

34	108.5	6.2	270	1	LPSC_RHIME	Q919m9 rhizobium m
35	107.5	6.1	297	1	Y060_MYCGE	P47306 mycoplasma
36	107.5	6.1	310	1	GTRB_BPP22	P57022 bacterioph
37	107	6.1	241	1	DPML_DROME	O9v1u7 drosophila
38	107	6.1	434	1	LPB1_LEITO	O05889 leishmania
39	106	6.0	256	1	Y157_MYCTU	O50459 mycobacteri
40	106	6.0	331	1	Y443_SYNY3	P74505 synechocyst
41	105	6.0	182	1	NODC_BRABL	P53417 bradyrhizob
42	103.5	5.9	243	1	YC22_METUA	O58619 methanococ
43	103	5.8	612	1	PAG3_CAEEL	P34678 caenorhabdi
44	102.5	5.8	424	1	NODC_RHILY	P04340 rhizobium l
45	101.5	5.8	334	1	ALG5_YEAST	P40350 saccharomyc

ALIGNMENTS

RESULT 1	YF78_HAEIN	STANDARD:	PRT:	323 AA.
ID	YF78_HAEIN			
AC	Q57287: 005077:			
DT	15-JUL-1998 (Rel. 36, Created)			
DT	15-JUL-1998 (Rel. 36, Last sequence update)			
DT	15-JUN-2002 (Rel. 41, Last annotation update)			
DE	Putative glycosyl transferase H11578 (EC 2.4.1.16).			
GN	H11578.			
OS	Haemophilus influenzae.			
OC	Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;			
OX	Haemophilus.			
NCBI_TaxID=727;				
RN	[1]			
RC	SEQUENCE FROM N.A.			
RC	STRAIN-Rd / KW20 / ATCC 51907:			
RX	MEDLINE-95350630; PubMed-7542800;			
RA	Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,			
RA	Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,			
RA	McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,			
RA	Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,			
RA	Weidman J.E., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,			
RA	Usterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,			
RA	Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,			
RA	Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,			
RA	Venter J.C.;			
RT	"Whole-genome random sequencing and assembly of Haemophilus influenzae			
RL	Rd.";			
CC	Science 269:496-512(1995).			
CC	-I- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.			
CC	-----			
CC	This SWISS-PROT entry is copyright. It is produced through a collaboration			
CC	between the Swiss Institute of Bioinformatics and the EMBL outstation -			
CC	the European Bioinformatics Institute. There are no restrictions on its			
CC	use by non-profit institutions as long as its content is in no way			
CC	modified and this statement is not removed. Usage by and for commercial			
CC	entities requires a license agreement (See http://www.isb-sib.ch/announce/			
CC	or send an email to license@sib-sib.ch).			
CC	-----			
DR	EMBL: U38832: AAC23227.1: -			
DR	TIGR: H11578: -			
DR	InterPro: IPR001173: Glycosyltransferase.			
DR	Pfam: PF00535: Glycosyltransferase_2: 1.			
KW	Hypothetical protein; Transferase; Glycosyltransferase;			
KW	Complete proteome.			
SO	SEQUENCE 323 AA; 37680 MW; 7C9C2681039A5B4 CRC64;			
Query Match	38.7%; Score 682.5; DB 1; Length 323;			
Best Local Similarity	45.9%; Pred. No. 3.7e-49;			
Matches 151; Conservative 54; Mismatches 107; Indels 17; Gaps 6;				
OY	3 PLVSVLICAYNAEKYFAQSLAVGQOTNRNDILLYVDOSTGCTALAHFPODGRRTI 62			
DB	5 PLVSVIVCAVNAEYTIDESTISITINOTYENLEITIVINDGSTFTTSHREISKLRKRI 64			
OY	63 ISNPNLGFIAISLNGIDELAKSGGGEYIARTDADDIASPGWIERKIVGEMEKDRSIAMG 122			

```

DB 65 ISKNYMLGFINSLTIGLGFCS-----GKPYAFMDADDIAKPSWTEIKIVLYLEKNDHTIANG 120
OY 123 ANLEVLISENNKNSVLAIAIRNGAINDKPTRHEDIYAVEFPNGPIHNTMIIRSVY-DGG 181
DB 121 SYLEIIVER-ECGIGISQYKRTSDIMKNPLHNDICAMLFYNDPIHNTMIIRANVYREHK 179
OY 182 LRFDPAYIAEDYKFEWYEGKLGRLAYPEALVKYRFHODQTSKYNLQORATWAKIKEE 241
DB 180 LIFNDYPAEDYKFEWYEGKLGRLAYPEALVKYRFHODQTSKYNLQORATWAKIKEE 239
OY 242 IRAGYWKAAGLAVGADCLNGLKSTAYALYERKALSGODIGCLRLPEYFLSLEKYSLT 301
DB 240 NITYYLNKIGIDIKV-----INSVSLLEYHYDKSNK---VLKSLYEWYMSLDKRYTIT 290
OY 302 DLDLFDLTVKMKLPAAPYOKRILKMLR 330
DB 291 SLHFI--KYHLELFDLKMOKLIKKEIR 317

```

RESULT 2

```

ID YAS7_METJA STANDARD; PRT; 290 AA.
AC Q58457;
DT 15-JUL-1998 (Rel. 36, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Putative glycosyl transferase MJ1057 (EC 2.-.-.-).
GN MJ1057.
OS Methanococcus jannaschii.
OC Archaea; Euryarchaeota; Methanococci; Methanococcales;
OC Methanocaldococcaceae; Methanocaldococcus.
OX NCBI_TaxID=2190;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=JAL-1 / DSM 2661 / ATCC 43067;
RX MEDLINE=9633799; PubMed=8688087;
RA Bult C.J., White O., Olsen G.J., Zhou L., Fleischmann R.D.,
RA Sutton G.G., Blake J.A., Fitzgerald L.M., Clayton R.A., Gocayne J.D.,
RA Kerlavage A.R., Dougherty B.A., Tomb J.-F., Adams M.D., Reich C.I.,
RA Overbeek R., Kirkness E.F., Weinstock K.G., Merrick J.M., Glodek A.,
RA Scott J.L., Geoghegan N.S.M., Weidman J.F., Fuhmann J.L., Nguyen D.,
RA Utterback T.R., Kelley J.M., Peterson J.D., Sadow P.W., Hanna M.C.,
RA Cotton M.D., Roberts K.M., Hurst M.A., Kaine B.P., Borodovsky M.,
RA Klenk H.-P., Fraser C.M., Smith H.O., Moese C.R., Venter J.C.;
RA "Complete genome sequence of the methanogenic archaeon, Methanococcus
RT jannaschii."
RL Science 273:1058-1073(1996).
CC -i SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC
DR EMBL: U67549; AAB9061.1; -
DR HSSP: P39621; IQGQ.
DR TIGR: MJ1057; -
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Hypothetical protein; Transferase; Complete proteome.
KW SEQUENCE 290 AA; 35099 MW; 3f6a1b221c420d74 CRC64;
SQ

```

```

Query Match 14.4%; Score 253.5; DB 1; Length 290;
Best Local Similarity 30.7%; Pred. No. 9,1e-14;
Matches 80; Conservative 51; Mismatches 101; Indels 29; Gaps 12;

```

```

OY 2 QPLVSVLCAVNAEKYPAQSLAAVVGQTVWRNLD-ILIVDDSGTSDGTPTAIARHFOQDGR 59
DB 7 KPLVSVVMAVNAEKYPAQSLAAVVGQTVWRNLD-ILIVDDSGTSDGTPTAIARHFOQDGR 66

```

```

OY 60 IRLSPNRLGFLASINIGLDELAKSGGEYIARDADDIASPGWTEIKYEGMEKDRSII 119
DB 67 IIFIKNERNLGAGASRNKAVN-IAR---GKIYALDADDIAPKRLREKPKYKMEENRDI 122
OY 120 ANCAMLEVLISENNKNSVLAIAIRNGAINDKPTRH--EDIVAVEFPNGPIHNTMIIRSV 177
DB 123 LIFSWMYFIDENGN--ILKEF-----KPEKYKFEIKKYF-FREHLTVHPSMWVSKS 171
OY 178 IDGGLRFPAYIAEDYKFEWYEGKLGRLAYPEALVKYRFHODQTSK-----YVL 229
DB 172 ILKKTKYDEKLRSODYDWIRCIANDYKFDIIEFLKYLPIPNRDNYSIRIKOKLYSY 231
OY 230 QORATWAKIKEE--RAGYWK 248
DB 232 YTLKTKWKNKKHPCNNVYFWK 252

```

RESULT 3

```

ID Y868_HAEIN STANDARD; PRT; 250 AA.
AC Q57022; P96336;
DT 15-JUL-1998 (Rel. 36, Created)
DT 15-JUL-1998 (Rel. 36, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase H10868 (EC 2.-.-.-).
GN H10868.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RX STRAIN=Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzgerald W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Utterback T.R., Hanna M.C., Nguyen D.T., Sauder D.W., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RL Science 269:496-512(1995).
CC -i SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

```

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch)
CC -----
CC
DR EMBL: U32768; AAC2526.1; -
DR HSSP: P39621; IQGQ.
DR TIGR: H10868; -
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
KW SEQUENCE 250 AA; 28915 MW; A5D8220129782B98 CRC64;
SQ

```

```

Query Match 11.8%; Score 209; DB 1; Length 250;
Best Local Similarity 22.4%; Pred. No. 3,6e-10;
Matches 68; Conservative 66; Mismatches 105; Indels 64; Gaps 10;

```

```

OY 3 PLVSVLCAVNAEKYPAQSLAAVVGQTVWRNLD-ILIVDDSGTSDGTPTAIARHFOQDGR 62
DB 4 PLISIMPVNAECYLNQILSLNOSYONIEILIDDDSGTSDKSIIRINNIIDKRVKL 63

```

Qy	63	ISNPRLGFIASLNGIDELAKSGGGEYIATDADDIASFGMIKRIYGEKEDSIIAMG	122
Db	64	FTFTPNQSPAAARNIGLEK-----AQSDYITFDSDDTIANDKLEKQJNFMLOJNLWTHG	119
Qy	123	AMLEVLSSENNKSVALAATARNGAITWDKPTRHEDIVAVEPPGPNPIHNNTIMRBSYIDGL	182
Db	120	NYAFCDLEGNOIKILVTT-----SKKIDYTLFLQJ-GNCFKIMTVIVERESIK-LI	166
Qy	183	REDPAVYIAHEOYKFMPEYAGKLGRLATYPEALVYKRFHQDQSSXVYNLQQRRTAMKIKEI	242
Db	167	RF-PNKKH-EDYAFFLDLCK-----EYAKOST	190
Qy	243	RAGYWKAAAGIAGVADCUNLYGLKST--AYALYEKALSGODIGCLR-----LFLYEFLS	294
Db	191	LYSHQASSFVRIGKVSYSNNKFKSAIMTFENIFYEKR---EKLGVVKSIIYFLLVAYNGEIK	247
Qy	295	LEK 297	
Db	248	YKK 250	

RESULT 4	
YS86_ANASP	
ID YS86_ANASP	STANDARD;
NO P33630.	PRT; 322 AA.

DT 01-AUG-1991 (Rel. 19, Created)
 DT 15-JUN-2002 (Rel. 41, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Purative glycosyl transferase alr2836 (EC 2.-.-.-).
 GN ALR2836.
 OS *Anabaena* sp. (strain PCC 7120).
 OC Bacteria; Cyanobacteria; Nostocales; Nostocaceae; Nostoc
 NCBI_taxid=103690;

RX MEDLINE-21595285; PubMed-11759848; Kurlitz T., Sasamoto S.,
 RA Kaneko T., Nakamura Y., Molk C.P.,
 RA Matsumoto M., Ishikawa A., Kawashima K., Kimura T.
 RA Kishida Y., Kohara M., Matsumoto M.,
 RA Nakazawa N., Shimpo S., Sugimoto M., Takazawa A., Yamada A.,
 RA Yasuda M., Rabata S.;
 RT "Complete genomic sequence of the filamentous nitrogen-fixing
 cyanobacterium *Anabaena* sp. strain PCC 7120.",
 RL DNA Res. 8:205-213(2001).

RN [2]
RP SEQUENCE OF 1-131 FROM N.A.
RX MEDLINE-90264305; PubMed-2111805;
RA Holland D., Wolk C.P.;
RT "Identification and characterization of beta, a gene that acts early
in the process of morphological differentiation of heterocysts."
RC J. Bacteriol. 173:3131-3137(1990).
CL -
CJ - SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.

 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation
 CC at the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@sib-ch.ch).

CC -----
DR EMBL; AP003591; BAB74535.1; -
DR EMBL; AF031959; AAC32401.1; ALT_INIT.
DR PIR; B35391; B35391.
DR InterPro: IPR001173; Glycos.transf.2.
DR Pfam: PF00535; Glycos.transf.2; 1.
KM Hypothetical protein; Glycosyltransferase
KM Complete proteome.
SQ SEQUENCE 322 AA; 36388 MW; 753C2FB59327D968 CRC64;

Query Match	11.1%;	Score 195.5;	DB 1;	Length 322
Best Local Similarity	23.7%;	Pred. NO. 6.5e-09;		

	Matches	74;	Conservative	57;	Mismatches	92;	Indels	89;	Gaps	15;
OY	5	VSVLICAVNAEYFFQSLAAVVGOTWRNDLLIVDDGSTDGPALAHFQEO-DGRIRII	63							
Dd	3	ISVIISNNVARYSLRAINSVLAQHSDIEIYVDGGSTDSNRDVTQTQQEQAPDKIKPI	62							
OY	64	SNPRULGTIASINIGLDELAKSGGEETIARTPADDIASGWLEKTIYGEMEKORSTIAMKA	123							
Dd	63	FQ-ANQGOGGGAFFNAGE----AATATEVAAFEDADVMKKRKLRIY-EVFQTSDVVGVMH	116							
OY	124	WLEVSEENKSVSVALAIARNGAIMDPTRHEIDIVY-----EPFGNPJHNNTMIRM	174							
Dd	117	HLDIT-DDNDRTIIDQASTQG-----PKLSDDLASYIILOTGNACKEP-----PTSQLAYR	164							
OY	175	RSVIDGLRFPPPAYTHADRYKFW-----YEAGKLGRLAYPEALVKYFRH-----	219							
Dd	165	REVLEKYVPIDPV---KWRIWAQGCILTYCAFLGKIIFTDLLENLAYYYIHGANNMMSAA	219							
OY	220	-----QD-----QTSSKY-----NLGOORT-----AMKIKEETR	243							
Dd	220	SATSDEAKSQAGILEMTQYINDFLVRIGYGARVDLSRNLYQRRTKYYORSOWDLRE---	276							
OY	244	AGYWKAAGIANG 255								
Dd	277	--WGISRLLIG 286								

RESULT	5
AMSE_ERWAM	
ID	AMSE_ERWAM
STANDARD;	PRT;
266	AA

AC Q46635;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Amyloryzarin biosynthesis glycosyl transferase amse (EC 2.-.-.-)
IN AMSE

GN
OS
OC
OC
OC
OX
OX
NN

Amoeb.
Ehrlichia amylovora.
Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae.
Ehrlichia.
NCBI_TaxID=552;
[1]

RN
 RP
 RC
 RX
 PA

L1
 SEQUENCE FROM N.A.
 STRAIN=EA1/79;
 MEDLINE=95319333; Pubmed=7596293;
 Budget B. Colfax

NA Digest: EcoRV, SmaI, XbaI.
RT Molecular analysis of the *ams* operon required for exopolysaccharide synthesis of *Erwinia amylovora*.
RL Mol. Microbiol. 15:917-933(1995).
CC - FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF AMYLOVORAN WHICH FUNCTIONS AS A VIRULENCE FACTOR.

CC -|- PATHWAY: Exopolysaccharide biosynthesis.
CC -|- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2
CC

 This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation at the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (see <http://www.isb-sib.ch/announce/sib.ch>) or send an email to license@sib-sb.ch.

CC -----
DR EMBL; X77921; CAA54886.1; -.
DR InterPro; IPR001173; Glycos_transf.2.
PIfam; PF00535; Glycos_transf.2; 1.
KW Exopolysaccharide synthesis; Transferrase; Glycosyltransferase
SQ SEQUENCE 266 AA; 30748 MW; 16326844210EB7B CRC64;

Query Match 11.0%; Score 194; DB 1; Length 266;
 Best Local Similarity 27.0%; Pred. No. 6.7e-09;
 Matches 80; Conservative 44; Mismatches 120; Indels 44; Gaps 14.

4 LVSVLICAYNAEK--YFAQSLAAVVGQTRNLDLIVDDGS-TDGTPAIARHFOEDGRI 60
 : ||||| : || : ||::|| : || : || :


```

0G Plasmid pSymb (megaplasmid 2).
0C Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
0C Rhizobiaceae; Sinothrixobium.
0X NCBI_TaxID=382;
0R [1]
0R RN SEQUENCE FROM N.A.
0R RP STRAIN=1021;
0C RC MEDLINE=94042869; PubMed=8226645;
0R RA Glucksmann M.A., Reuber T.L., Walker G.C.;
0R RT "Family of glycosyl transferases needed for the synthesis of
0R succinoglycan by Rhizobium meliloti.";
0R RL J. Bacteriol. 175:7033-7044(1993).
0R [2]
0R RN SEQUENCE FROM N.A.
0R RP STRAIN=RCR2011 / SU47;
0C RC MEDLINE=94067019; PubMed=8246891;
0R RA Becker A., Kleickmann A., Keller M., Arnold W., Pehler A.;
0R RT "Identification and analysis of the Rhizobium meliloti exoAMONP genes
0R involved in exopolysaccharide biosynthesis and mapping of promoters
0R located on the exoHKLAMONP fragment.";
0R RL Mol. Gen. Genet. 241:367-379(1993).
0R [3]
0R RN SEQUENCE FROM N.A.
0R RP STRAIN=1021;
0C RC MEDLINE=21396508; PubMed=11481431;
0R RA Fianan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
0R Vorhoefer F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
0R Golding B., Pehler A.;
0R RT "The complete sequence of the 1,683-kb pSymb megaplasmid from the N2-
0R fixing endosymbiont Sinothrixobium meliloti.";
0C RC Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
0C CC -1- FUNCTION: GLYCOSYLTRANSFERASE REQUIRED FOR THE SYNTHESIS OF
0C CC SUCCINOGLYCAN (EPS I). NEEDED FOR THE ADDITION OF THE SECOND SUGAR
0C CC (GLUCOSE). CATALYZES THE FORMATION OF A BETA-1,3 LINKAGE WITH THE
0C CC GALACTOSE LIPID CARRIER.
0C CC -1- PATHWAY: Exopolysaccharide biosynthesis.
0C CC -1- SUBCELLULAR LOCATION: Integral membrane protein.
0C CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
0C CC -----
0C CC This SWISS-PROT entry is copyright. It is produced through a collaboration
0C CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
0C CC the European Bioinformatics Institute. There are no restrictions on its
0C CC use by non-profit institutions as long as its content is in no way
0C CC modified and this statement is not removed. Usage by and for commercial
0C CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
0C CC or send an email to license@sib-sib.ch).
0C CC -----
0C CC
0R DR EMBL, L20758; AAA16046.1; -.
0R CC
0R DR EMBL, Z22636; CAAB0345.1; -.
0R CC
0R DR EMBL, AL603645; CAC49482.1; -.
0R CC
0R DR PIR, S37027; S37027.
0R CC
0R DR PIR, S39956; S39956.
0R CC
0R DR InterPro: IPR001173; Glycosyltransf_2.
0R CC
0R DR Pfam: PF00535; Glycosyltransf_2; 1.
0R CC
0R RW Transferase; Glycosyltransferase; Exopolysaccharide synthesis;
0R KW Transmembrane; Plasmid; Complete proteome.
0R FT TRANSMEM 116 136 POTENTIAL.
0R FT TRANSMEM 260 280 POTENTIAL.
0R FT TRANSMEM 299 319 POTENTIAL.
0R SO SEQUENCE 330 AA; 36167 MW; 6A6A919F52602F1 CRC64;
0R
0R Query Match 9.3%; Score 164.5; DB 1; Length 330;
0R Best Local Similarity 24.6%; Pred. NO.2.4e-06;
0R Matches 67; Conservative 39; Mismatches 105; Indels 61; Gaps 11;
0R
0Y 32 NLDIIVDGSSTGTPATARHFOEDGRIRIISNPNLGFISLNIIGDELAKSGGGEYI 91
0Y I::I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
0Y Db 39 NARVVIADGSGSDGRREIARLRATEDPVLFDNKKRIQ-SAAVRAVAL--GAGSDVL 95
0Y
0Y 92 ARTDADDIASPGWIEKIVGEMEKDRSIITAMGAMLEVLSE--NNKSYLAIAIRNGAIV 147
0Y I I I I I I I I I I I I I I I I I I I I I I I I I I I I I I
0Y Db 96 IRIDAHGTYPDDYCERLY-----EDALATAGDGSVVAMQGTGFTFOKATPAFONSKLG 149

```

OY 148 DKRFHEDIVAFPPGNGNIHNNTMIIMRSVIDGLRDPATIAHEBKYFEAGKLG---
Db 150 NGGSKHR--TGAVGHAAEHGHALMKLEFAKVGG--VDSESHNEDAELLYRUSKAQYRI
OY 205 -----RLAYPEALLVKKFRHQ-----DOTSSKNYLDOOR-----
Db 207 WMTKTSTSVYYTPRAKIADPLFWMQRYFGYGGRKANLTKLRAMPDGLRMPLAVAPITAFALL 266

OY 234 --TWAKIKEEIRIAGTYWKAAGAIVAGACLNLNGCL 263
Db 267 AIVNW--MAVPVGVMAAA-----CLGYGV 289

RESULT 10
SPSA_BACSU STANDARD; PRT; 256 AA.
ID SPSA_BACSU STANDARD; PRT; 256 AA.
AC P39621;
DT 01-FEB-1995 (Rel. 31, Created)
DT 01-FEB-1995 (Rel. 31, Last sequence update)
DT 15-JUN-2002 (Rel. 41, last annotation update)
DE Spore coat polysaccharide biosynthesis protein spsa.
GN SPSA OR IPA 63D.
OS Bacteria subtilis.
OC Bacillus Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_Taxid=1423;

RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95040037; PubMed=7934828;
RA Azevedo V., Bertero M.G., Bessieres P., Bolotin A., Borcherdt S.,
RA Borriss R., Boursier L., Brans A., Braun M., Bridgell S.C., Bron S.,
RA Bouillet S., Brunshi C.V., Caldwell B., Capiano V., Carter N.M.,
RA Choi S.K., Codani J.U., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denizot F., Devine K.M., Dusterhoft A., Erlich S.D., Emerson P.T.,
RA Enlian K.D., Erlington J.Y., Fabret C., Ferrari E., Fougeret D.,
RA Fritz C., Fujita Y., Fujita Y., Funai S., Galizzi A., Galleron N.,
RA Ghim S.-Y., Glaser P., Goffeau A., Goldlithy E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Hatech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullio M.F., Itaya M., Jones L.,
RA Joris B., Katamata D., Kasahara Y., Klaerr-Bianchini M., Klein C.,
Ra Kobayashi Y., Koehler P., Koningsstein G., Krogh S., Kunano M.,
RA Kusita K., Lapidas A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Manuel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Mostl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Outega B., Park S.H.,
RA Parro V., Pohl T.M., Portetelle D., Potworilik S., Prescott A.M.,
RA Piessens E., Puljic P., Punelle B., Rapoport G., Rey M., Reynolds S.,
RA Rieger M., Riyvota C., Rocha E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Seliguchi J., Sekowska A., Seror S.J., Seror P., Shin B.S., Soldo B.,
RA Socokin A., Tacconi E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Tanakoshi A., Tanaka T., Terpsira P., Togroni A.,
RA Tosato V., Uchiyama S., Vandendobel M., Vanlier F., Vassarotti A.,
RA Viriri A., Mambutt R., Wedler E., Wedler H., Weitzenegeer T.,
RA Winter P., Wiipit A., Yamamoto H., Yanane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshioka H.F., Zmistein E., Yoshioka H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249 -256(1997).
RP X-RAY CRYSTALLOGRAPHY (1.5 ANGSTROMS).

RA MEDLINE-99280694; PubMed-10350455;
RT Charnock S.J.; Davies G.J.;
RT "Structure of the nucleotide-diphospho-sugar transferase, SpssA from
RT Bacillus subtilis, in native and nucleotide-complexed forms.";
RL Biochemistry 38:6380-6385(1999).
CC -1- FUNCTION: GLYCOSYLTRANSFERASE IMPLICATED IN THE SYNTHESIS OF THE
CC SPORE COAT.
CC -1- PATHWAY: Spore coat polysaccharide biosynthesis.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC -----
CC EMBL: X73124; CAAS1619.1; -
CC EMBL: 299123; CAB15817.1; -
CC PDB: 1OG8; 21-APR-00.
CC PDB: 1OG8; 04-MAY-00.
CC PDB: 1OGS; 04-MAY-00.
CC Subtilist; Bg10609; spsa.
CC InterPro: IPR001173; Glycos_transf_2.
CC Pfam: PF00535; Glycos_transf_2; 1.
CC Transferase; Glycosyltransferase; 3D-structure; Complete proteome.
FT DISULFID 155 243
SQ SEQUENCE 256 AA; 30184 MW; C20EA9627F5D536B CRC64;

Query Match 9.1%; Score 160.5; DB 1; Length 256;
Best Local Similarity 22.1%; Pred. No. 3.7e-06;
Matches 60; Conservative 61; Mismatches 98; Indels 53; Gaps 12;

QY 3 PLVSVLTICAYNAKEKYFAOASLAAVGOTNRNLDILYDSDSTGCTPAIAHFOQDGRIR- 61
DB 2 PAVSYALMSTYNSDYAKSISSILSTQSFDFELIMDDNSNETLWVIRPFL-NDNRKVF 60
QY 62 -----IISNRNLGFIASLNGIDELAKSGGEYIARTDADIASPGMIKIVGEM- 113
DB 61 YQSDISGVKEPTEKRYALINQAI-EMAE---GEYITATDQNIYMPRLKMWELDT 116
QY 114 -KDRSIANGAELEVLSEENKSVLAIRNGA--IWDKPTRHEDIVAVFPFGNPIHNT 170
DB 117 HEKAVIYASAKTYHLNE--NRDIYKEIVRPAQVYWNAPC-----AIDHCS 161
QY 171 IMRRSVIDG-----GLRFD--PAYIHADYKFWYAGKLGRLAYP-----E 211
DB 162 VMHRSVLEKVEKFGSYWDESPAFYRIDARFWRVNHF--YPPYLDDELNLNYITDQ 219
QY 212 ALVKYRFHODQTSKYNLQOORTAWRIKEIR 243
DB 220 SIHFQLELEKNEFVRNLPPQNRCLRESLTK 251

RESULT 11
EXOO_RHIME STANDARD; PRT; 348 AA.
ID EXOO_RHIME
AC P33697;
DT 01-FEB-1994 (Rel. 28, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Suctionglycan biosynthesis protein exoo (EC 2.---).
GN EXOO OR RB1084 OR SMB20959.
OS Rhizobium meliloti (Sinorhizobium meliloti).
OC Plasmid pSymB (megaplasmid 2).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Rhizobiaceae; Sinorhizobium.
OX NCBI_TaxID=382;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE=94042869; PubMed=8226645;

RA Glucksmann M.A.; Reuber T.L.; Walker G.C.;
RT "Family of glycosyl transferases needed for the synthesis of
RT suctionglycan by Rhizobium meliloti";
RL J. Bacteriol. 175:7033-7044(1993).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=RCR2011 / 5047;
RX MEDLINE-94067019; PubMed-8246891;
RA Becker A., Kleckmann A., Keller M., Arnold W., Puehler A.;
RT "Identification and analysis of the Rhizobium meliloti exoAMONP genes
RT involved in exopolysaccharide biosynthesis and mapping of promoters
RT located on the exoKIAMONP fragment";
RL Mol. Gen. Genet. 241:367-379(1993).
RN [3]
RP SEQUENCE FROM N.A.
RC STRAIN=1021;
RX MEDLINE-21396508; PubMed-11481431;
RA Finan T.M., Weidner S., Wong K., Buhrmester J., Chain P.,
RA Vorholter F.J., Hernandez-Lucas I., Becker A., Cowie A., Gouzy J.,
RA Golding B., Puehler A.;
RT "The complete sequence of the 1,683-kb pSymB megaplasmid from the N2-
RT fixing endosymbiont Sinorhizobium meliloti";
RL Proc. Natl. Acad. Sci. U.S.A. 98:9889-9894(2001).
CC -1- FUNCTION: GLYCOSYLTRANSFERASE REQUIRED FOR THE SYNTHESIS OF
CC (GLUCOSE), CATALYZES THE FORMATION OF A BETA-1,6 LINKAGE BETWEEN
CC THE FOURTH AND FIFTH SUGAR.
CC -1- PATHWAY: Exopolysaccharide biosynthesis.
CC -1- SUBCELLULAR LOCATION: Cytoplasmic.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -1- CAUTION: Ref.1 sequence differs from that shown due to a
CC frameshift in position 208.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation-
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@sib.ch).
CC -----
CC EMBL: L20758; AAL16044.1; ALT_FRAME.
CC EMBL: 222636; CAAB0347.1; -
CC EMBL: AL603645; CAC49484.1; -
CC PIR: C49348; C49348.
CC PIR: S37029; S37029.
CC PIR: S39958; S39958.
CC InterPro: IPR001173; Glycos_transf_2.
CC Pfam: PF00535; Glycos_transf_2; 1.
CC Transferase; Glycosyltransferase; Exopolysaccharide synthesis;
CC Plasmid; Complete proteome.
FT DOMAIN 38 45
FT POLY-VAL.
SQ SEQUENCE 348 AA; 38131 MW; 770B43782F785579 CRC64;

Query Match 8.8%; Score 155; DB 1; Length 348;
Best Local Similarity 25.4%; Pred. No. 1.6e-05;
Matches 60; Conservative 40; Mismatches 104; Indels 32; Gaps 8;

QY 3 PLVSVLTICAYNAKEKYFAOASLAAVGOTNRNLDILYDSDSTGCTPAIAHFOQDGRIR 62
DB 10 PDTEFVAAYNSADTIVRAIESALAEQGVTEVVVDDCSAATPALVAAL--PDPVRVL 67
QY 63 ISNPNLGFASLNGIDELAKSGGEYIARTDADIASPGMIKIVGEMKDRSIIM- 121
DB 68 IALDRNRGCGARNAGIG-----AARGRWLAIVDSDDTVPDRIRRRRIERADAAGAIADV 123
QY 122 -----GAMLEVLSEEN-----NKSVLAIARNGAIIWDKPTRHEDIVAVFPFGNPIHNN 169
DB 124 NLDVVSLDGRSLRMFSEAEIARLPQTLPAFIESVLF--RSEHN-----FGYMKRIFE- 175
QY 170 TIMRRSVIDGIRDPAYIHADYKFWYEGAKL-GRLAYPEALVKYFHHODQTS 224
DB 176 ----RRFLENQDLRFDEALRIGEDYTLILASALACGRCACVPSAGYIYHIREGIS 227

```
RESULT 12
YCDQ_ECOLI
ID YCDQ_ECOLI STANDARD: PRT; 441 AA.
AC P75905;
DT 01-NOV-1997 (rel. 35, Created)
DT 01-NOV-1997 (rel. 35, Last sequence update)
DT 16-OCT-2001 (rel. 40, Last annotation update)
DE Hypothetical protein ycdq.
GN YCDQ OR B1022.
OS Escherichia coli.
OC Bacteria; Proteobacteria; gamma subdivision; Enterobacteriaceae;
OC Escherichia.
OX NCBI_TaxID=562;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=K12 / MG1655;
RX MEDLINE=97426617; PubMed=9278503;
RA Blattner F.R., Plunkett G. III, Bloch C.A., Perna N.T., Burland V.,
RA Riley M., Collado-Vides J., Glasner J.D., Rode C.K., Mayhew G.F.,
RA Gregor J., Davis N.W., Kirkpatrick H.A., Goeden M.A., Rose D.J.,
RA Mau B., Shao Y.;
RT "The complete genome sequence of Escherichia coli K-12.";
RL Science 277:1453-1474(1997).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=K12;
RX MEDLINE=97061202; PubMed=8905232;
RA Oshima T., Alba H., Baba T., Fujita K., Hayashi K., Honjo A.,
RA Ikemoto K., Inada T., Itoh T., Kajihara M., Kanai K., Kashimoto K.,
RA Kimura S., Kitegawa M., Makino K., Masuda S., Miki T., Mizobuchi K.,
RA Mori H., Motomura K., Nakamura Y., Nishimoto H., Nishio Y., Saito N.,
RA Sasaki G., Seki Y., Tagami H., Takemoto K., Wada C., Yamamoto Y.,
RA Yano M., Horinouchi T.;
RT "A 718-kb DNA sequence of the Escherichia coli K-12 genome
corresponding to the 12.7-28.0 min region on the linkage map.";
RL DNA Res. 3:137-155(1996).
CC -1 SUBCELLULAR LOCATION: Integral membrane protein (Potential).
CC -1 SIMILARITY: TO Y_PESRIS HMS LOCUS PROTEIN HMSR AND TO
CC S_EPIDERMIDIS ICA4.
-----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sdb.ch/announce/
CC or send an email to license@sdb.ch).
-----
CC EMBL; AE000204; AAC74107.1; -
CC EMBL; D90739; BAA35803.1; -
CC DR Ecocore: EGI1363; ycdq.
CC DR InterPro: IPR001173; Glycos_transf_2.
CC DR Pfam: PF00535; Glycos_transf_2; 1.
CC KW Hypothetical protein; Transmembrane; Complete proteome.
CC FT TRANSMEM 5 25 POTENTIAL.
CC FT TRANSMEM 33 53 POTENTIAL.
CC FT TRANSMEM 331 351 POTENTIAL.
CC FT TRANSMEM 363 383 POTENTIAL.
CC FT TRANSMEM 393 415 POTENTIAL.
CC FT SEQUENCE 441 AA; 50765 MW; 04F5A5D72FEBAB CRC64;
QY
QY 3 PLVSVLICAVNAEKYQAQSIAAVVGOTWRNLDLIVDGSSTGCPALARHFOEDGRIRI 62
QY 75 PSISIIIPCEKNEKVEETIHAAQRYENIEIVAVNDGSTDKTRAILDMAAQIPHLRV 134
QY 63 ISNPRRLGFTASINIGLDELAKSGGGEYIARTADD-----IASPQWIKIYGEV 113
```

```
DB 135 IHLAQNQGAIALKTG---AAAKSEYLCIDGDALLDQDAAYIYEPMLYNPRGAVT 190
QY 114 KDRSIITAMGAWLEVLSEENKSVLAIAIRNGAIWDPKTRHEDIVAVPEFGNPITHNTMIM 173
DB 191 GNPRIKTRSLVKIKYQGESSIIGLIKRTQRIYGNVFTYSSVIAAF----- 237
QY 174 RRSVIDGGLRFDPAYIHAEDYKFEYAGKLGRLAYYPEAL-----VKYRFHODQTSSKY 227
DB 238 RRSAA-----LAEVGYSMDMITEDIDISMKLQNMQNTIFY 272
QY 228 NLOQRRTAMKIKKEIRAGYWK 248
DB 273 -EPRALCWITMBETLKGIMK 291
RESULT 13
YWDF_BACSU
ID YWDF_BACSU STANDARD: PRT; 268 AA.
AC P39614;
DT 01-FEB-1995 (rel. 31, Created)
DT 01-FEB-1995 (rel. 31, Last sequence update)
DT 15-JUN-2002 (rel. 41, Last annotation update)
DE Putative glycosyl transferase ywdf (EC 2.-.-.-).
GN YWDF OR IPA-56D.
OS Bacillus subtilis.
OC Bacteria; Firmicutes; Bacillales; Bacillaceae; Bacillus.
OX NCBI_TaxID=1423;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=95020537; PubMed=7934828;
RA Glaser P., Kunst F., Arnaud M., Coudat M.P., Gonzales W.,
RA Hullo M.F., Ionescu M., Lubochinsky B., Marcelino L., Mosser I.,
RA Presecan E., Tanescu M., Schneider E., Schweizer J., Vertes A.,
RA Rapoport G., Danchin A.;
RT "Bacillus subtilis genome project: cloning and sequencing of the 97
kb region from 325 degrees to 333 degrees.";
RL Mol. Microbiol. 10:371-384(1993).
[2]
RP SEQUENCE FROM N.A.
RC STRAIN=168;
RX MEDLINE=98044033; PubMed=9384377;
RA Kunst F., Ogasawara N., Moszer I., Albertini A.M., Alloni G.,
RA Azevedo V., Baetjer M.G., Bessieres P., Bolotin A., Borcherdt S.,
RA Bortis R., Boursier L., Brans A., Braun M., Brignell S.C., Bron S.,
RA Brouillet S., Bruschi C.V., Caldwell B., Capuano V., Carter N.M.,
RA Choi S.K., Codani J.J., Connerton I.F., Cummings N.J., Daniel R.A.,
RA Denicot F., Devine K.M., Dusterhoft A., Ehrlich S.D., Emerson P.T.,
RA Entlan K.D., Errington J., Fabret C., Ferrati E., Foulger D.,
RA Fritz C., Fujita M., Fujita Y., Funa S., Galizzi A., Galleron N.,
RA Ghim S.Y., Glaser P., Goffeau A., Golightly E.J., Grandi G.,
RA Giuseppe G., Guy B.J., Haga K., Halech J., Harwood C.R., Henaut A.,
RA Hilbert H., Holtsappel S., Hosono S., Hullo M.F., Itaya M., Jones L.,
RA Joris B., Karamata D., Kasahara Y., Klier-Blanchard M., Klein C.,
RA Kobayashi Y., Koeltter P., Koningsstein G., Krogh S., Kumano M.,
RA Kurita K., Lapidus A., Lardinois S., Lauber J., Lazarevic V.,
RA Lee S.M., Levine A., Liu H., Masuda S., Mauel C., Medigue C.,
RA Medina N., Mellado R.P., Mizuno M., Moestl D., Nakai S., Noback M.,
RA Noone D., O'Reilly M., Ogawa K., Ogihara A., Oudega B., Park S.H.,
RA Parro V., Pohl T.M., Portetle D., Porwollik S., Prescott A.M.,
RA Presecan E., Pujic P., Purnelle B., Rapoport G., Ray M., Reynolds S.,
RA Rieger M., Rivolta C., Roche E., Roche B., Rose M., Sadale Y.,
RA Sato T., Scanlan E., Schleich S., Schroeter R., Scoffone F.,
RA Sekiguchi J., Sekowski A., Serot S.J., Serrot P., Shin B.S., Soldo B.,
RA Sorokin A., Taccioni E., Takagi T., Takahashi H., Takemaru K.,
RA Takeuchi M., Yamakoshi A., Tanaka T., Terpstra P., Tognoni A.,
RA Tosato V., Uchiyama S., Vandenbol M., Vannier F., Vassartoli A.,
RA Viari A., Wambutt R., Wedler E., Wedler H., Weltzenegger T.,
RA Winters P., Wipat A., Yamamoto H., Yamane K., Yasunoto K., Yata K.,
RA Yoshida K., Yoshikawa H.F., Zimstein E., Yoshikawa H., Danchin A.;
RT "The complete genome sequence of the Gram-positive bacterium Bacillus
subtilis.";
RL Nature 390:249-256(1997).
```

```

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: X73124; CA51612.1; -.
DR EMBL: Z99123; CAB15824.1; -.
DR PIR: S39711; S39711.
DR Subtilist; B610602; ywGF.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
DR Hypothetical protein; Transferase; Glycosyltransferase;
KW Complete proteome.
SQ SEQUENCE 268 AA; 30616 MW; DD8428F7016EC9B3 CRC64;

Query Match
Best Local Similarity 25.4%; Score 147; DB 1; Length 268;
Matches 65; Conservative 45; Mismatches 120; Indels 26; Gaps 9;

OY 5 VSVLICAVNAKRYAQAISLAAVGQ-TWRNLDILYDOSTGTCTPAIARHFOBODGRIRIS 64
DB 3 ISIIYVTRNRIPALCELLSISROTLMPEYELIIVDAGESEVYVYALY---PELPIAVIN 59
OY 65 NPNRNGFIASINIGDELAKSGGGEYIARTDADIASPGWTEKIVGEMEKRSITIAMAM 124
DB 60 LEKNSGVHAARNAGV---KEASGGCIIMCDDDFEYFGHEKMAKETE-TADVFHSDA- 113
OY 125 LEVLSENNKSVLAIAIRNGAIWDKPTRHEDIVAF---PEGNPIHNNTIMMRSVIDG 180
DB 114 -EIVSEFEKNGTRPVSKR--LFAVTADYED-MRFSYTYVSGS-----MYRFLHDE 162
OY 181 GLRDPATIAHEDKFWYEAQKGLRLAYPPALYKYPHQ---DOTSSKYNLQQRRTANKI 238
DB 163 IGYPDADVHNWDMDFYLRRAKDYRVKPCASVYAFSDAGDNOSADLGAKRKQYDLRL 222
OY 239 KEIRPAGYKAGIATV 254
DB 223 SEKHGELPTKRRAY 238

RESULT 14
YF20 MYCTU STANDARD; PRT; 346 AA.
AC 050587;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical protein Rv1520.
GN Rv1520 OR MT1570 OR MTCY195.08C.
OS Mycobacterium tuberculosis.
OC Bacteria; Actinobacteria; Actinobacteria (class); Actinobacteridae;
OC Actinomycetales; Corynebacterineae; Mycobacteriaceae; Mycobacterium.
OX NCBI_TaxID=1773;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=H37Rv.
RA MEDLINE=98295987; Pubmed=9634230;
RA Cole S.T., Brosch R., Parkhill J., Garnier T., Churcher C., Harris D.,
RA Gordon S.V., Eiglmeier K., Gas S., Barry C.E. III, Tekala F.,
RA Badcock K., Basham D., Brown D., Chillingworth T., Connor R.,
RA Davies R., Devlin K., Felwell T., Gentles S., Hamlin N., Holroyd S.,
RA Honesy T., Jagels K., Krogh A., McLean J., Moule S., Murphy L.,
RA Oliver S., Osborne K., Quail M.A., Rajandream M.A., Rogers J.,
RA Rutter S., Seeger K., Skelton S., Squares S., Squares R.,
RA Sulston J.E., Taylor K., Whitehead S., Barrell B.G.;
RT "Deciphering the biology of Mycobacterium tuberculosis from the
RT complete genome sequence.";
RL Nature 393:537-544(1998).

```

```

RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=CDC 1551 / Oshkosh;
RA Fleisemann R.D., Alland D., Eisen J.A., Carpenter L., White O.,
RA Peterson J., Deboy R., Dodson R., Gwinn M.L., Haft D., Hickey E.,
RA Kolony J.F., Nelson W.C., Umayam L.A., Ermolaeva M.D., Salzberg S.L.,
RA Delcher A., Utterback T., Weidman J., Khouri H., Gill J., Mikula A.,
RA Bisbal W.;
RT "Whole genome comparison of Mycobacterium tuberculosis clinical and
RT laboratory strains";
RL Submitted (Apr-2001) to the EMBL/GenBank/DBJ databases.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Z77826; CAB01396.1; -.
DR EMBL: AE007024; AAK45837.1; ALT_INIT.
DR TIGR: MT1570; -.
DR Tuberculist; Rv1520; -.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
DR Hypothetical protein; Complete proteome.
SQ SEQUENCE 346 AA; 39578 MW; 22EDA0495E0EFC6A CRC64;

Query Match
Best Local Similarity 22.2%; Score 142; DB 1; Length 346;
Matches 60; Conservative 41; Mismatches 109; Indels 60; Gaps 9;

OY 5 VSVLICAVNAKRYAQAISLAAVGQ-TWRNLDILYDOSTGTCTPAIARHFOBODGRIRI 62
DB 1 MSIVSISTNQEYIREALDGFQAORTPEFVEVITADASTATPRITIGEVARYPOLFRP 60
OY 63 ISNPNLGFISINIGDELAKSGGGEYIARTDADDI-ASPGWTEKIVGEMEKRSITAM 121
DB 61 ILRQTNIGVHANFK---DVLSARGEYALCEBGDDVTWTDLKLISKQYKIDRRPETHVC 116
OY 122 GAMLEVLSENNK-----SVLAIAIRNGAIWDKPTRHEDIVAFPGNPIN 168
DB 117 FHPVRIYIEDGAKSEFPPLSMRDLSDVALLARNF-----IQT 155
OY 169 NTMTMRSVIDGRLFD--PAYIAHEDKFWYEAQKGLRLAYPPALYKYPHQDQTSK 226
DB 156 NSVYRRP-----SYDDIPANVMPIDWYLHVRHNAVGGELMPEYAVYRRH----- 203
OY 227 YNLQQRRTANKIKEIRPAGYKAGIATV 256
DB 204 -----AHGIWHSATYTRDRKRFETRGHGMAA 228

RESULT 15
YG95 HAEIN STANDARD; PRT; 267 AA.
AC 048215; 005081;
DT 16-OCT-2001 (Rel. 40, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Putative glycosyl transferase H11695 (EC 2.-.-.-).
GN H11695.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellales;
OC Haemophilus.
OX NCBI_TaxID=727;
RN 11
RP SEQUENCE FROM N.A.
RC STRAIN=A2;
RA McLaughlin R., Abu Kwaik Y., Young R., Spinoia S., Apicella M.;
RT "Characterization and sequence of the lsg locus from Haemophilus
RT influenzae.";

```

```
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KM20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McEnaney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uterback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.;
RT "Whole-genome random sequencing and assembly of Haemophilus influenzae
Rd.";
RL Science 269:496-512(1995).
CC -1 SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 2.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
CC DR EMBL: M94855; AAC24983.1; -.
CC DR EMBL: U32842; AAC23341.1; -.
CC DR TIGR: H11693; -.
CC DR InterPro: IPR001173; Glycosyltransferase_2.
CC DR Pfam: PF00535; Glycosyltransferase_2; 1.
CC KW Hypothetical protein; Transferase; Glycosyltransferase;
CC Complete proteome.
CC FT CONFLICT 26 26 V -> G (IN REF. 1).
CC FT CONFLICT 46 46 D -> E (IN REF. 1).
CC FT CONFLICT 49 49 F -> S (IN REF. 1).
CC SQ SEQUENCE 267 AA; 30770 MW; A2P1A0532737D8C3 CRC64;

Query Match 8.0%; Score 140.5; DB 1; Length 267;
Best local similarity 25.5%; Pred. No. 0.00018;
Matches 63; Conservative 45; Mismatches 114; Indels 25; Gaps 9;

OY 6 SVLICAYNAE--KYPFASIAAVVGQWNRMLDILIVDGSSTGTPAIAAHFOEDGR--IR 61
DB 4 SVLMSLYIKENQFLRECPESLVAQTRADEIVLVEDGV--TPDLEFVYVEFEETKPLK 61
OY 62 IISNPRNIGFIASLNIUGDELAKSGGGEYIARTDADIASPGWIEKIVGEMEKRSIIAM 121
DB 62 IYKLPQNRGLGRALNEGILHC---DYDWVFRMDTDIDICVPRFEKOVAFIEQHPESTIF 117
OY 122 GAWLEVLSEPNKSVLAIAIRNGAIDMKPTREHEDIVAVFPFGNPIHNNMTIMRS-VIDG 180
DB 118 GGQIAEFGKNVNDIV-----I::II::II::II::II::II::II::II::II::II 169
OY 181 GLRPDPAYIHADYKFWYEAGKLG-RLAYVYPALVKYRFHODQTSKYNLQQRRTAMKI- 238
DB 170 GGVED---LQGEDYIWMIKIVAGLYMANLPDILIVARVNGVSRRGVNGAKAEMRLF 225
OY 239 KEIRIAG 245
DB 226 KIKYRIAG 232
```

Search completed: December 2, 2002, 11:56:29
Job time : 10.2019 secs

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 28.9477 Seconds

(without alignments)
2398.732 Million cell updates/sec

Title: US-10-007-267-5

Perfect score: 1764

Sequence: 1 MQLPVSLICAVNAEKYFAQ.....APQYRKILKMLRPMWKYSY 337

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 671580 seqs, 206047115 residues

Total number of hits satisfying chosen parameters: 671580

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : SPTREMBL_21.*
1: sp-archaea:*
2: sp-bacteria:*
3: sp-fungi:*
4: sp-human:*
5: sp-invertebrate:*
6: sp-mammal:*
7: sp-mhc:*
8: sp-organelle:*
9: sp-phage:*
10: sp-plant:*
11: sp-rodent:*
12: sp-virus:*
13: sp-vertebrate:*
14: sp-unclassified:*
15: sp-virus:*
16: sp-bacteriap:*
17: sp-archaeap:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1764	100.0	337	2	050949
2	1079	61.2	348	2	09EVD5
3	1074.5	60.9	333	2	051115
4	1069.5	60.6	349	16	09JX06
5	1055	59.8	346	2	093PS0
6	1049	59.5	348	2	050946
7	1048	59.4	346	2	050946
8	1045	59.2	362	2	050951
9	1039	58.9	346	16	09JW66
10	953	54.0	321	2	09LME9
11	942.5	53.4	322	2	09LRF0
12	712.5	40.4	337	16	09CIR9
13	588	33.3	221	2	093EK6
14	433	24.5	172	2	09RCM9
15	415	23.5	149	2	P96946
16	392	22.2	337	2	034234

17	391	22.2	337	2	P74947	P74947 vibrio chol
18	346	19.6	367	16	092V61	092V61 rhizobium m
19	330.5	18.7	368	16	098JH2	098JH2 rhizobium l
20	282.5	16.0	333	16	097H38	097H38 clostridium
21	275	15.6	117	2	P96944	P96944 neisseria m
22	272	15.4	133	2	093EK8	093EK8 neisseria m
23	265.5	15.1	314	2	09LA88	09LA88 aeromonas h
24	263.5	14.9	336	16	097H39	097H39 clostridium
25	258.5	14.7	340	2	087159	087159 vibrio chol
26	258	14.6	333	16	097H40	097H40 clostridium
27	249	14.1	271	16	09K6L5	09K6L5 bacillus ha
28	244.5	13.9	343	16	08YUP7	08YUP7 anabaena sp
29	243	13.8	298	17	09U216	09U216 pyrococcus
30	237.5	13.5	288	2	056869	056869 yeastia en
31	237	13.4	80	2	09K2R3	09K2R3 neisseria g
32	236	13.4	257	2	093C09	093C09 shigella bo
33	234	13.3	337	16	08YD6	08YD6 anabaena sp
34	231	13.1	278	16	P71054	P71054 bacillus su
35	228.5	13.0	298	17	08U2R3	08U2R3 pyrococcus
36	226.5	12.8	301	16	097P78	097P78 streptococ
37	225.5	12.8	318	16	08YSI4	08YSI4 anabaena sp
38	225	12.8	318	16	08YSI3	08YSI3 anabaena sp
39	225	12.8	334	17	058167	058167 pyrococcus
40	224	12.7	732	16	09K6S0	09K6S0 bacillus ha
41	223.5	12.7	311	2	082874	082874 streptococ
42	223	12.6	321	16	08YSM2	08YSM2 anabaena sp
43	223	12.6	333	16	08YWS0	08YWS0 anabaena sp
44	221.5	12.6	324	2	08YLB0	08YLB0 streptococ
45	221	12.5	344	16	P71057	P71057 bacillus su

ALIGNMENTS

RESULT 1	ID	Q50949	PRELIMINARY;	PRT;	337 AA.
AC	Q50949;	01-NOV-1996 (TREMBLrel. 01, Created)			
DT	01-NOV-1996 (TREMBLrel. 01, Last sequence update)				
DE	01-JUN-2001 (TREMBLrel. 17, Last annotation update)				
DE	Glycosyl transferase.				
GN	LGTD.				
OS	Neisseria gonorrhoeae.				
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.				
OX	NCBI_TaxID=485;				
RN	[1]				
RP	SEQUENCE FROM N.A.				
RC	STRAIN=F62;				
RX	MEDLINE=95053752; PubMed=7964493;				
RA	Gotschlich E.C.;				
RT	"Genetic locus for the biosynthesis of the variable portion of				
RT	Neisseria gonorrhoeae lipooligosaccharide.";				
RL	J. Exp. Med. 180:2181-2190(1994).				
DR	EMBL; U14554; AAA68012.1;				
DR	Interpro: IPR001173; Glycos. transfr_2.				
DR	Pfam: PF00535; Glycos. transfr_2.1.				
KW	Transferase.				
SQ	SEQUENCE 337 AA; 38393 MW; 5AE21AA7SCAA732E CRC64;				
Query Match	100.0%; Score 1764; DB 2; Length 337;				
Best Local Similarity	100.0%; Pred. No. 4.8e-137;				
Matches 337; Conservative	0; Mismatches 0; Indels 0; Gaps 0;				
QY	1 MQLPVSLICAVNAEKYFAQSLAAVVGQTRNDLILYDGSSTDGTPATARFQEQDGR 60				
DB	1 MQLPVSLICAVNAEKYFAQSLAAVVGQTRNDLILYDGSSTDGTPATARFQEQDGR 60				
QY	61 RIISNPRNIGFTASLNTIGDELAKSGGGEYIARTADDDIASGWIETKGEKDRSITA 120				
DB	61 RIISNPRNIGFTASLNTIGDELAKSGGGEYIARTADDDIASGWIETKGEKDRSITA 120				
QY	121 MCAMLEVLSEENNKSVLAIAIRNGAIWDPKTRHEDIIVAVFPFGNPINHTMTMRSDVDG 180				


```

Db 121 MGAMLEVLSEENKSVLAIAARNGAIWDKPTRHEDIVAVFPFGNPJHNTMIRRSYIDG 180
OY 181 GLRPDPAVIAHAEYKFWYKGLGRLAYYPEALVYKRFHODQSSKYNLQOORTAKIKE 240
Db 181 GLRPDPAVIAHAEYKFWYKGLGRLAYYPEALVYKRFHODQSSKYNLQOORTAKIKE 240
OY 241 EIRAGYWKAAAGIAGVADCLNYGLKSTAYALYKALSGODICLRLFLYEYFLSLEKYSL 300
Db 241 EIRAGYWKAAAGIAGVADCLNYGLKSTAYALYKALSGODICLRLFLYEYFLSLEKYSL 300
OY 301 TDLLDFLTDVRMKRLFAAPQYRKILKMLRPWKYRST 337
Db 301 TDLLDFLTDVRMKRLFAAPQYRKILKMLRPWKYRST 337

RESULT 2
O9EVD5 PRELIMINARY: PRT: 348 AA.
AC O9EVD5;
DT 01-MAR-2001 (TREMblrel. 16, Created)
DT 01-MAR-2001 (TREMblrel. 16, Last sequence update)
DT 01-JUN-2002 (TREMblrel. 21, Last annotation update)
DE N-acetylglucosamine transferase.
GN LGTA.
OS Neisseria subflava.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=28449;
RN [1]
RP SEQUENCE FROM N.A.
RA MEDLINE=21142520; PubMed=11208792;
RA Arking D., Tong Y., Stein D.C.;
RT "Analysis of lipooligosaccharide biosynthesis in the Neisseriaceae.";
RL J. Bacteriol. 183:934-941(2001).
DR EMBL: AF240672; AAC09764.1; -.
DR HSRP: P39621; I0G0.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
KW Transferase.
SQ SEQUENCE 348 AA; 40676 MW; BC9E313E9BC0BF41 CRC64;

Query Match 61.2%; Score 1079; DB 2; Length 348;
Best Local Similarity 66.1%; Pred. No. 1.2e-80;
Matches 218; Conservative 29; Mismatches 83; Indels 0; Gaps 0;

OY 1 MOPVLSVLCAYNAEKYFQSLAAVVGOTWRNDLILYDGSIDGTPATARRHEDDGR 60
Db 17 LQPLVSLICAYVEKTFQSLAAVVGOTWRNDLILYDGSIDGTPATARRHEDDGR 76
OY 61 RIISNPNLGFIALNIGLDELAKSGGGEYIARTDADDIASPQMIKIVGEMEKDSIIA 120
Db 77 RIIAQRNSGLIPLNIGLDELAKSGGGEYIARTDADDIAPMIKIVGEMEKDSIIA 136
OY 121 MGAMLEVLSEENKSVLAIAARNGAIWDKPTRHEDIVAVFPFGNPJHNTMIRRSYIDG 180
Db 137 MGAMLEVLSEENKSVLAIAARNGAIWDKPTRHEDIVAVFPFGNPJHNTMIRRSYIDG 196
OY 181 GLRPDPAVIAHAEYKFWYKGLGRLAYYPEALVYKRFHODQSSKYNLQOORTAKIKE 240
Db 197 GLRPDPAVIAHAEYKFWYKGLGRLAYYPEALVYKRFHODQSSKYNLQOORTAKIKE 256
OY 241 EIRAGYWKAAAGIAGVADCLNYGLKSTAYALYKALSGODICLRLFLYEYFLSLEKYSL 300
Db 257 TARNDLFQSMGKTRPDSLEYRQIKAVAYELLEKHLPEEDFERARRLFYQCFKRTDPLPA 316
OY 301 TDLLDFLTDVRMKRLFAAPQYRKILKMLR 330
Db 317 GAWLDEPAADGRMRRLFTLRQYFGILHRLK 346

RESULT 3
O51115 PRELIMINARY: PRT: 333 AA.

```

```

AC O51115;
DT 01-NOV-1996 (TREMblrel. 01, Created)
DT 01-NOV-1996 (TREMblrel. 01, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Glycosyl transferase.
GN LGTA.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=487;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=MC58;
RC MEDLINE=96414473; PubMed=8817494;
RA Jennings M.P., Hood D., Peak I.R.A., Virji M., Moxon E.R.;
RT "Molecular analysis of a locus for the biosynthesis and phase-variable expression of the lacto-N-neotetraose terminal lipopolysaccharide structure in Neisseria meningitidis.";
RL Mol. Microbiol. 18:729-740(1995).
DR EMBL: U25839; AAC44084.1; -.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
SQ SEQUENCE 333 AA; 38563 MW; 49D8F6CE375387BF CRC64;

Query Match 60.9%; Score 1074.5; DB 2; Length 333;
Best Local Similarity 65.9%; Pred. No. 2.7e-80;
Matches 218; Conservative 29; Mismatches 83; Indels 1; Gaps 1;

OY 1 MOPVLSVLCAYNAEKYFQSLAAVVGOTWRNDLILYDGSIDGTPATARRHEDDGR 60
Db 1 MOPVLSVLCAYVEKTFQSLAAVVGOTWRNDLILYDGSIDGTPATARRHEDDGR 60
OY 61 RIISNPNLGFIALNIGLDELAKSGGGEYIARTDADDIASPQMIKIVGEMEKDSII 119
Db 61 RIIAQRNSGLIPLNIGLDELAKSGGGEYIARTDADDIAPMIKIVGEMEKDSII 120
OY 120 AMGAMLEVLSEENKSVLAIAARNGAIWDKPTRHEDIVAVFPFGNPJHNTMIRRSYID 179
Db 121 AMGAMLEVLSEENKSVLAIAARNGAIWDKPTRHEDIVAVFPFGNPJHNTMIRRSYID 180
OY 180 GGLRPDPAVIAHAEYKFWYKGLGRLAYYPEALVYKRFHODQSSKYNLQOORTAKIK 239
Db 181 GGLRPDPAVIAHAEYKFWYKGLGRLAYYPEALVYKRFHODQSSKYNLQOORTAKIK 240
OY 240 EIRAGYWKAAAGIAGVADCLNYGLKSTAYALYKALSGODICLRLFLYEYFLSLEKYS 299
Db 241 KTRANDLFQSMGKTRPDSLEYRQIKAVAYELLEKHLPEEDFERARRLFYQCFKRTDPL 300
OY 300 LNDLDFLTDVRMKRLFAAPQYRKILKMLR 330
Db 301 AGAWLDEPAADGRMRRLFTLRQYFGILHRLK 331

RESULT 4
O9JX06 PRELIMINARY: PRT: 349 AA.
AC O9JX06;
DT 01-OCT-2000 (TREMblrel. 15, Created)
DT 01-OCT-2000 (TREMblrel. 15, Last sequence update)
DT 01-DEC-2001 (TREMblrel. 19, Last annotation update)
DE Lacto-N-neotetraose diosynthesis glycosyl transferase LgtA.
GN NMB1929.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN=MC58 / SEROGRUP B;
RC MEDLINE=20117575; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E., Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J., Nelson W.C., Gwin M.L., Deboy R., Peterson J.D., Hickey E.K., Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A., Mason T., Ciecko A., Parksey D.S., Blair E., Ciltone H., Clark E.B.,

```



```

SQ SEQUENCE 346 AA; 40158 MW; 95C89CBA01FFA9 CRC64;
Query Match 58.9%; Score 1039; DB 16; Length 346;
Best Local Similarity 63.3%; Pred. No. 2.4e-77;
Matches 209; Conservative 35; Mismatches 86; Indels 0; Gaps 0;
QY 1 M0PLVSVLICANNKRYFAQSLAAVVGOTWRNLDILIVDDGSTDGTPTAARHFOEDGR1 60
DB 1 M0PLVSVLICANNKRYFAQSLAAVVGOTWRNLDILIVDDGSTDGTPTAARHFOEDGR1 60
QY 61 RIISNPNRLGFIASLNLGIDELAKSGGGEYIARTDADDIASFGWIKTEKIVGEMEKDRS1TA 120
DB 61 RIISNPNRLGFIASLNLGIDELAKSGGGEYIARTDADDIASFGWIKTEKIVGEMEKDRS1TA 120
QY 121 M0AMLEVEISEENKSVLAIAARNGA1MPKTRHEDIVAVFPFGNPIHNNTMTMRSSVIDG 180
DB 121 M0AMLEVEISEENKSVLAIAARNGA1MPKTRHEDIVAVFPFGNPIHNNTMTMRSSVIDG 180
QY 181 GLRDPAYIHAEDYKFWYEGAKGLRAYVPEALVYRPHODOTSSKYNLQ0RRTAMKIKE 240
DB 181 GLRYNTERDMADYQFWYDVSKLGRLAYVPEALVYRLHANVSSKYSIRQHEINAGIOIX 240
QY 241 EIRAGYWKRAAGIAGVADCLNIGLKSTAYALYKALSGODICLRLFLYEYLSLEKISL 300
DB 241 TARNDFL0SMGFKTRPDSLEYROIKAVALLEKHLPEDEFERARFLYQCFKRTDTPPA 300
QY 301 TDLDFLTRVVRKLEFAAPQYRKILKKMR 330
DB 301 GAMLDEADGKKRRLLFTMRQYFGLIHLRIK 330

RESULT 10
Q9L8E9 PRELIMINARY; PRT; 321 AA.
AC Q9L8E9;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LGTA.
GN LGTA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=SS2;
RA Baltnazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
RT "Mutations in the 1gt operon influence serum-resistance in
RT gonococci."
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF208059; AAF25877.1; -.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
SQ SEQUENCE 321 AA; 37052 MW; 6ACADA9A3CB738FA CRC64;
Query Match 54.0%; Score 953; DB 2; Length 321;
Best Local Similarity 62.6%; Pred. No. 2.6e-70;
Matches 191; Conservative 34; Mismatches 80; Indels 0; Gaps 0;
QY 26 VQOTWRNLDILIVDDGSTDGTPTAARHFOEDGRIRIISNPNRLGFIASLNLGIDELAKS 85
DB 1 M0QOTWRNLDILIVDDGSTDGTPTAARHFOEDGRIRIISNPNRLGFIASLNLGIDELAKS 85
QY 86 GGGEYIARTDADDIASFGWIKTEKIVGEMEKDRSIIAMGAWLEVLSEENKSVLAIAARNGA 145
DB 61 GGGEYIARTDADDIASFGWIKTEKIVGEMEKDRSIIAMGAWLEVLSEENKSVLAIAARNGA 145
QY 146 IWDKPTRHEDIIVAVFPFGNPIHNNTMTMRSSVIDGLRFPDPAIYIHAEDYKFWYEGAKLG 205
DB 121 IWDKPTRHEDIIVAVFPFGNPIHNNTMTMRSSVIDGLRFPDPAIYIHAEDYKFWYEGAKLG 205
QY 206 LAYYPEALVYKRYRLHANVSSKSHVROHEIAQGIQKTARNDFL0SMGFKTRPDSLEYROT 240
DB 301 LYRLIK 306

RESULT 11
Q9L8F0 PRELIMINARY; PRT; 322 AA.
AC Q9L8F0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LGTA.
GN LGTA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA899;
RA Baltnazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
RT "Mutations in the 1gt operon influence serum-resistance in
RT gonococci."
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF208058; AAF25876.1; -.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
SQ SEQUENCE 322 AA; 37109 MW; 88FECA5AD41DC2B7 CRC64;
Query Match 53.4%; Score 942.5; DB 2; Length 322;
Best Local Similarity 62.4%; Pred. No. 1.9e-69;
Matches 191; Conservative 34; Mismatches 80; Indels 1; Gaps 1;
QY 26 VQOTWRNLDILIVDDGSTDGTPTAARHFOEDGRIRIISNPNRLGFIASLNLGIDELAKS 85
DB 1 M0QOTWRNLDILIVDDGSTDGTPTAARHFOEDGRIRIISNPNRLGFIASLNLGIDELAKS 85
QY 86 -GGGEYIARTDADDIASFGWIKTEKIVGEMEKDRSIIAMGAWLEVLSEENKSVLAIAARNGA 144
DB 61 GGGEYIARTDADDIASFGWIKTEKIVGEMEKDRSIIAMGAWLEVLSEENKSVLAIAARNGA 144
QY 145 A1WMDPTRHEDIIVAVFPFGNPIHNNTMTMRSSVIDGLRFPDPAIYIHAEDYKFWYEGAKLG 204
DB 121 A1WMDPTRHEDIIVAVFPFGNPIHNNTMTMRSSVIDGLRFPDPAIYIHAEDYKFWYEGAKLG 204
QY 205 RLAYYPEALVYKRYRLHANVSSKSHVROHEIAQGIQKTARNDFL0SMGFKTRPDSLEYROT 240
DB 181 RLAYYPEALVYKRYRLHANVSSKSHVROHEIAQGIQKTARNDFL0SMGFKTRPDSLEYROT 240
QY 241 KAAAYELPEKDLPEDEFERARFLYQCFKRTDTPPSGAWLDEADGRMRRLFTLRYQYGI 300
DB 301 LYRLIK 306

RESULT 12
Q9CLR9 PRELIMINARY; PRT; 337 AA.
AC Q9CLR9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical protein Pml140.
GN Pml140.
```

```

DB 181 LAYYPEALVYKRYRLHANVSSKSHVROHEIAQGIQKTARNDFL0SMGFKTRPDSLEYROT 240
QY 266 SNAAYLYEKALSGODICLRLFLYEYFLSLEKYSILTDLDLFDIVMRKLEFAAPQYRKIL 325
DB 241 AAAYELPEKDLPEDEFERARFLYQCFKRTDTPPSGAWLDEADGRMRRLFTLRYQYGI 300
QY 326 KKMLR 330
DB 301 YRLIK 305

RESULT 1
Q9L8F0 PRELIMINARY; PRT; 322 AA.
AC Q9L8F0;
DT 01-OCT-2000 (TREMBLrel. 15, Created)
DT 01-OCT-2000 (TREMBLrel. 15, Last sequence update)
DT 01-JUN-2001 (TREMBLrel. 17, Last annotation update)
DE LGTA.
GN LGTA.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FA899;
RA Baltnazar J.T., Shafer W.M., Stephens D.S., Martin L.E.;
RT "Mutations in the 1gt operon influence serum-resistance in
RT gonococci."
RL Submitted (NOV-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF208058; AAF25876.1; -.
DR InterPro: IPR001173; Glycos_transf_2.
DR Pfam: PF00535; Glycos_transf_2; 1.
SQ SEQUENCE 322 AA; 37109 MW; 88FECA5AD41DC2B7 CRC64;
Query Match 53.4%; Score 942.5; DB 2; Length 322;
Best Local Similarity 62.4%; Pred. No. 1.9e-69;
Matches 191; Conservative 34; Mismatches 80; Indels 1; Gaps 1;
QY 26 VQOTWRNLDILIVDDGSTDGTPTAARHFOEDGRIRIISNPNRLGFIASLNLGIDELAKS 85
DB 1 M0QOTWRNLDILIVDDGSTDGTPTAARHFOEDGRIRIISNPNRLGFIASLNLGIDELAKS 85
QY 86 -GGGEYIARTDADDIASFGWIKTEKIVGEMEKDRSIIAMGAWLEVLSEENKSVLAIAARNGA 144
DB 61 GGGEYIARTDADDIASFGWIKTEKIVGEMEKDRSIIAMGAWLEVLSEENKSVLAIAARNGA 144
QY 145 A1WMDPTRHEDIIVAVFPFGNPIHNNTMTMRSSVIDGLRFPDPAIYIHAEDYKFWYEGAKLG 204
DB 121 A1WMDPTRHEDIIVAVFPFGNPIHNNTMTMRSSVIDGLRFPDPAIYIHAEDYKFWYEGAKLG 204
QY 205 RLAYYPEALVYKRYRLHANVSSKSHVROHEIAQGIQKTARNDFL0SMGFKTRPDSLEYROT 240
DB 181 RLAYYPEALVYKRYRLHANVSSKSHVROHEIAQGIQKTARNDFL0SMGFKTRPDSLEYROT 240
QY 241 KAAAYELPEKDLPEDEFERARFLYQCFKRTDTPPSGAWLDEADGRMRRLFTLRYQYGI 300
DB 301 LYRLIK 306

RESULT 12
Q9CLR9 PRELIMINARY; PRT; 337 AA.
AC Q9CLR9;
DT 01-JUN-2001 (TREMBLrel. 17, Created)
DT 01-JUN-2001 (TREMBLrel. 17, Last sequence update)
DT 01-DEC-2001 (TREMBLrel. 19, Last annotation update)
DE Hypothetical protein Pml140.
GN Pml140.
```

OS Pasteurella multocida.
 OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
 OC Pasteurella.
 OX NCBI_TaxID=747;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=PM70;
 RX MEDLINE=21145866; PubMed=11248100;
 RA May B.J., Zhang Q., Li L.L., Paustian M.L., Paustian T.S., Kapur V.;
 RT "Complete genomic sequence of Pasteurella multocida PM70.";
 RL Proc. Natl. Acad. Sci. U.S.A. 98:3460-3465(2001).
 DR EMBL: AE006155; AAK03224.1;
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2.1.
 KW Hypothetical protein; Complete proteome.
 SQ SEQUENCE 337 AA; 39267 MW; 8639BCFB5F700DB4 CRC64;

Query Match 40.4%; Score 712.5; DB 16; Length 337;
 Best Local Similarity 46.2%; Pred. No. 1.7e-50;
 Matches 156; Conservative 56; Mismatches 115; Indels 11; Gaps 5;

OY 2 OPLVSLICAVNAEKYFAOSLAAYVGQTRNDILIVDDGSTGTPAIAHFEODGRIR 61
 DB 6 OPLVSLICAVNAADKIIECIDAILMOTYKNLEIVVNDGSTDTTSLKHYFGKDPRIK 65
 OY 62 IISPNRLGFIASLNLGDELAASKSGGEYIARTDADDIASPCWIEKIVGEMEKDRSIAM 121
 DB 66 IINENKKGFIASLNLGSIASLNGDYLRARDADDTIKPEWIEKILGVMLSHPQIIAM 121
 OY 122 GAMLEVLSENNKSVLAIAIRNGATMDKPTREDDIYAVPEFGPNINNTYINRSYI-DC 180
 DB 122 GSYLTRLSDDGNSNLANYEHGDEWRNPLSHREIVEAMLFNPNHNSMIVSTVEREH 181
 OY 181 GLRFDPAYTHAEDKFKWYEGKLGRLAYPEALVKYRFHODOTSSYXNQOVRTAKIKE 240
 DB 182 GLRFDPAYTHEDYQFWLEYSKRGELANTPESLVYIRLHNTQTSLHNKYQNMAMKIRK 241
 OY 241 EIRAGYAKAGI--AVGADCLNGLKSTAYALYERALKSGODIGCLRLFYEFSLSEKY 298
 DB 242 RAINVYLQDGLVYHRLGEDIFPHDIFIQAEL--ASLSLNDLCIIRLKYDCLSLVDN 298
 OY 299 SLTDLLDFLTRVMKRLFAAPQYRKILKMLRPWKYRS 336
 DB 299 KLINILYFLDRK-NNSYFNKKOKIKIKIRIRPKYKES 335

RESULT 13
 OY93EK6 PRELIMINARY; PRT; 221 AA.
 AC Q93EK6;
 DT 01-DEC-2001 (TREMBLrel. 19, Created)
 DT 01-DEC-2001 (TREMBLrel. 19, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE LGTA.
 GN LGTA.
 OS Neisseria meningitidis.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=487;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=M978;
 RX MEDLINE=21467954; PubMed=11583844;
 RA Zhu P., Klutch M.J., Tsai C.-M.;
 RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
 Expression in Two 18-Immunotype Strains of Neisseria meningitidis.";
 RL FEMS Microbiol. Lett. 203:173-177(2001).
 DR EMBL: AF355193; AAL12840.1;
 SQ SEQUENCE 221 AA; 26478 MW; D1P97099B1F6055F CRC64;

Query Match 33.3%; Score 588; DB 2; Length 221;
 Best Local Similarity 55.3%; Pred. No. 1.7e-40;
 Matches 121; Conservative 25; Mismatches 73; Indels 0; Gaps 0;

OY 112 MEKDSIIMGAMLEVLSENNKSVLAIAIRNGATMDKPTREDDIYAVPEFGPNINNTM 171
 DB 1 MEKDSIIMGAMLEVLSENNKSVLAIAIRNGATMDKPTREDDIYAVPEFGPNINNTM 60
 OY 172 IMRSYIDGGLPDPAYTHAEDKFKWYEGKLGRLAYPEALVKYRFHODOTSSYXNQO 231
 DB 61 IMRSYIDGGLPDPAYTHAEDKFKWYEGKLGRLAYPEALVKYRFHODOTSSYXNQO 120
 OY 232 RRTAMKKEIRIAGYKAGIAGVADCLNGLKSTAYALYERALKSGODIGCLRLFYEX 291
 DB 121 HELAGSIQKTANNDLFQSGNGFTFRDSDLEYROIKAAYELKHLPEDEFEARARFLVOC 180
 OY 292 FLSLEKYSTLTDLDFLTRVMKRLFAAPQYRKILKMLR 330
 DB 181 FKRTDTLPAGAWLDFEADBGMRRLFTLRQYFGILHRLK 219

RESULT 14
 OY9RG9 PRELIMINARY; PRT; 172 AA.
 AC O9RG9;
 DT 01-MAY-2000 (TREMBLrel. 13, Created)
 DT 01-MAY-2000 (TREMBLrel. 13, last sequence update)
 DT 01-DEC-2001 (TREMBLrel. 19, last annotation update)
 DE Glycosyltransferase.
 GN LGTD.
 OS Neisseria gonorrhoeae.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=485;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=1291;
 RX MEDLINE=20305049; PubMed=10844691;
 RA Harey H.A., Porat N., Campbell C.A., Jennings M., Gibson B.W.,
 RA Phillips N.J., Apicella M.A., Baile M.S.;
 RT "Genococcal lipooligosaccharide is a ligand for the asialoglycoprotein
 receptor on human sperm."
 RL Mol. Microbiol. 36:1059-1070(2000).
 DR EMBL: AF121135; AAF14362.1;
 DR InterPro: IPR001173; Glycos_transf_2.
 DR Pfam: PF00535; Glycos_transf_2.1.
 KW Transferase.
 SQ SEQUENCE 172 AA; 18872 MW; 7D2737434EDEFCC4 CRC64;

Query Match 24.5%; Score 433; DB 2; Length 172;
 Best Local Similarity 96.6%; Pred. No. 6.6e-28;
 Matches 85; Conservative 1; Mismatches 2; Indels 0; Gaps 0;

OY 1 MOPVSLICAVNAEKYFAOSLAAYVGQTRNDILIVDDGSTGTPAIAHFEODGRIR 60
 DB 1 MOPVSLICAVNAEKYFAOSLAAYVGQTRNDILIVDDGSTGTPAIAHFEODGRIR 60
 OY 61 RIISPNRLGFIASLNLGDELAASKSGG 88
 DB 61 RIISPNRLGFIASLNLGDELAASKSGG 88

RESULT 15
 P96946 PRELIMINARY; PRT; 149 AA.
 AC P96946;
 DT 01-MAY-1997 (TREMBLrel. 03, Created)
 DT 01-MAY-1997 (TREMBLrel. 03, last sequence update)
 DT 01-JUN-2002 (TREMBLrel. 21, last annotation update)
 DE Glycosyl transferase.
 GN LGTD.
 OS Neisseria meningitidis.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=487;
 RN [1]
 RP SEQUENCE FROM N.A.
 RC STRAIN=126E;
 RX MEDLINE=20055626; PubMed=10589709;

RA Jennings M.P., Strikanta Y.N., Moxon E.R., Kramer M., Poolman J.T.,
 RA Kuipers B., van der Ley P.;
 RT "The genetic basis of the phase variation repertoire of
 RT lipopolysaccharide immunotypes in *Neisseria meningitidis*.";
 RL Microbiology 145:3013-3021(1999).
 DR EMBL; U65788; AAB48386.1; -.
 DR HSSP; P39621; 10GQ.
 DR InterPro; IPR001173; Glycos_transf_2.
 DR Pfam; PF00535; Glycos_transf_2; 1.
 KW Transferrase.
 SQ SEQUENCE 149 AA; 16667 MW; 68A52EB7DF6B552A CRC64;

Query Match 23.5%; Score 415; DB 2; Length 149;
 Best Local Similarity 83.3%; Pred. No. 1.6e-26;
 Matches 85; Conservative 3; Mismatches 12; Indels 2; Gaps 1;

QY 1 MQLVSVLICAVNAEKYFAQSIAAVVGOTWRNLDTLIYDDGSTDGTPTAIRHFOEDGRI 60
 DB 1 MQLVSVLICAVNAEKYFAQSIAAVVGOTWRNLDTLIYDDGSTDGTPTAIRHFOEDGRI 60
 QY 61 RIISNPNLGFATSLNIGDELAKSGGGEYIARTDADDIASP 102
 DB 61 RIISNPNLGFATSLNIGDELAKSGGGEYIARTDADDIASP 100

Search completed: December 2, 2002, 12:02:18
 Job time : 29.9477 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 29.0623 Seconds
(without alignments)
1283.801 Million cell updates/sec

Title: US-10-007-267-6

Perfect score: 1440
Sequence: 1 MGNHVISLASAERRAHAD.....ERQRAELEKVGGRVILFK 280

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 908470 seqs, 133250620 residues

Total number of hits satisfying chosen parameters: 908470

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database :

1: A_Geneseq_101002.*
2: /SID2/gcgdata/geneseq/geneseq-emb1/AA1980.DAT.*
3: /SID2/gcgdata/geneseq/geneseq-emb1/AA1981.DAT.*
4: /SID2/gcgdata/geneseq/geneseq-emb1/AA1982.DAT.*
5: /SID2/gcgdata/geneseq/geneseq-emb1/AA1983.DAT.*
6: /SID2/gcgdata/geneseq/geneseq-emb1/AA1984.DAT.*
7: /SID2/gcgdata/geneseq/geneseq-emb1/AA1985.DAT.*
8: /SID2/gcgdata/geneseq/geneseq-emb1/AA1986.DAT.*
9: /SID2/gcgdata/geneseq/geneseq-emb1/AA1987.DAT.*
10: /SID2/gcgdata/geneseq/geneseq-emb1/AA1988.DAT.*
11: /SID2/gcgdata/geneseq/geneseq-emb1/AA1990.DAT.*
12: /SID2/gcgdata/geneseq/geneseq-emb1/AA1991.DAT.*
13: /SID2/gcgdata/geneseq/geneseq-emb1/AA1992.DAT.*
14: /SID2/gcgdata/geneseq/geneseq-emb1/AA1993.DAT.*
15: /SID2/gcgdata/geneseq/geneseq-emb1/AA1994.DAT.*
16: /SID2/gcgdata/geneseq/geneseq-emb1/AA1995.DAT.*
17: /SID2/gcgdata/geneseq/geneseq-emb1/AA1996.DAT.*
18: /SID2/gcgdata/geneseq/geneseq-emb1/AA1997.DAT.*
19: /SID2/gcgdata/geneseq/geneseq-emb1/AA1998.DAT.*
20: /SID2/gcgdata/geneseq/geneseq-emb1/AA1999.DAT.*
21: /SID2/gcgdata/geneseq/geneseq-emb1/AA2000.DAT.*
22: /SID2/gcgdata/geneseq/geneseq-emb1/AA2001.DAT.*
23: /SID2/gcgdata/geneseq/geneseq-emb1/AA2002.DAT.*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1440	100.0	280	17	AA91315
2	1425	99.0	280	18	AA906580
3	1036.5	72.0	279	17	AA91312
4	1007	69.9	275	20	AA922156
5	233	16.2	50	20	AA969331
6	144	10.0	1128	22	AB862795
7	129.5	9.0	423	23	AB89453
8	129.5	9.0	517	21	AA841790
9	129.5	9.0	517	22	AA93202
10	129.5	9.0	517	22	AA93206

11	129.5	9.0	578	22	AA681257	Human AFP protein
12	129.5	9.0	579	20	AA930812	Human secreted pro
13	129.5	9.0	595	21	AA812121	Hydrophobic domain
14	129.5	9.0	636	23	AB942016	Human ovarian anti
15	127	8.8	521	21	AA841900	Human ORFX ORF1664
16	127	8.8	622	21	AA93427	Human polypeptide,
17	127	8.8	622	22	AA88452	Human membrane or
18	127	8.8	622	23	AB943477	Human secreted pro
19	109	7.6	739	22	AB811417	Human secreted pro
20	108.5	7.5	168	22	AB820950	Novel human diagno
21	106.5	7.4	201	18	AA920604	H. pylori cytoplas
22	104	7.2	273	22	AA860317	Helicobacter pylor
23	103	7.2	279	18	AA921023	H. pylori cytoplas
24	99	6.9	394	22	AA924257	Aquifex aspartate
25	97	6.7	273	22	AA860321	Helicobacter pylor
26	97	6.7	444	23	AA976669	Helicobacter pylor
27	95	6.6	273	20	AA93263	H. pylori beta-1,4
28	94	6.5	767	22	AB611534	Novel human diagno
29	94	6.5	4010	22	AB861520	Drosophila melanog
30	94	6.3	389	22	AB82459	Caenorhabditis ele
31	90.5	6.3	991	16	AA880096	Black widow spider
32	90.5	6.3	1214	16	AA880097	Black widow spider
33	88.5	6.1	161	23	AA815943	Worm C38H2-2/141-3
34	87.5	6.1	657	22	AA933784	Staphylococcus aur
35	87.5	6.1	662	21	AA808632	Amino acid sequenc
36	87.5	6.1	662	21	AA936547	Staphylococcus aur
37	87.5	6.1	1129	19	AA977286	Bovine differentia
38	87.5	6.1	1343	22	AA933259	Novel human secret
39	87	6.0	1674	19	AA881109	Human BAK1-alpha p
40	86	6.0	291	23	AB849964	Listeria monocytog
41	86	6.0	526	22	AB892861	Human protein sequ
42	85.5	5.9	653	23	AB847763	Listeria monocytog
43	85	5.9	439	22	AA930597	Novel human secret
44	85	5.9	753	20	AA937247	Protein involved i
45	85	5.9	762	22	AA935926	Helicobacter pylor

ALIGNMENTS

RESULT 1	AA91315	standard; Protein: 280 AA.
AA91315		
XX	AA91315	
AC	AA91315	
XX	09-JUL-1996	(first entry)
DT		
XX		
DE	N. gonorrhoeae glycosyltransferase LgtE.	
XX		
KW	Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;	
KW	vaccine.	
XX		
OS	Neisseria gonorrhoeae strain F62.	
XX		
PN	W09610086-A1.	
XX		
PD	04-APR-1996.	
XX		
PE	25-SEP-1995;	95WO-US12317.
XX		
PR	26-SEP-1994;	94US-0312387.
XX		
PA	(UTR) UNIV ROCKEFELLER.	
XX		
PI	Gotschlich EC;	
XX		
DR	WPI: 1996-200924/20.	
XX		
DR	N-PSDB: AAT14061.	
XX		
FT	Nucleic acids encoding glycosyl transferase(s) - used in the	
PT	diagnosis of infection with Neisseria and for the biosynthesis of	
PT	oligo:saccharide(s)	

xx Claim 12; Fig 2f; 81bp; English.
 PS 5 glycosyltransferases (AAR91311-15) are products of the lgt locus
 CC (AAT14061) of *Neisseria gonorrhoeae* strain F62. Glycosyltransferase
 CC LgtE (AAR91315) can be obtained by expression of the lgtE coding
 CC sequence in recombinant host cells. A method for adding Gal
 CC beta1-4 to GlcNAc or Glc comprises contacting a reaction mixture
 CC containing activated Gal to an acceptor moiety comprising a GlcNAc or
 CC Glc residue in the presence of LgtE. Oligosaccharides can be produced
 CC that, when attached to non-toxic lipids, are useful for *Neisseria*
 CC vaccine prepns. Blood group core oligosaccharides, and mimics of
 CC lacto-N-neotetraose, gangliosides and saccharide portions of
 CC globoglycolipids can also be produced using the enzymes.
 CC
 SQ Sequence 280 AA:
 Query Match 100.0%; Score 1440; DB 17; Length 280;
 Best Local Similarity 100.0%; Pred. No. 6,4e-148;
 Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
 QY 1 MONHYSLSAERRAHIAADTFGSRGIPFOFPDAMPSERLQAMAEIVPGISAHPYLSG 60
 DB 1 MONHYSLSAERRAHIAADTFGSRGIPFOFPDAMPSERLQAMAEIVPGISAHPYLSG 60
 QY 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLGGAQOFLAEDTWLEERFDKDSAFIVR 120
 DB 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLGGAQOFLAEDTWLEERFDKDSAFIVR 120
 QY 121 LETMFAKYIVRPDKYLVNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPPERIKA 180
 DB 121 LETMFAKYIVRPDKYLVNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPPERIKA 180
 QY 181 VDIAMFTFFDEKGMPIYOVSPALCTQELHYAKFLSONSMGSDLEKDEGRGRRRSLSK 240
 DB 181 VDIAMFTFFDEKGMPIYOVSPALCTQELHYAKFLSONSMGSDLEKDEGRGRRRSLSK 240
 QY 241 VFEDLKRALGKRGREKKRMRQROAELEKYGRVYILFK 280
 DB 241 VFEDLKRALGKRGREKKRMRQROAELEKYGRVYILFK 280
 RESULT 2
 AAM06580
 ID AAM06580 standard; Protein; 280 AA.
 AC AAM06580;
 XX
 DT 21-MAR-1997 (first entry)
 DE Lipo-oligosaccharide gene-encoded protein.
 XX
 KW Polyglycosyltransferase; N-acetylglucosaminyl transferase;
 KW N-acetylglactosaminyl transferase; lipo-oligosaccharide.
 XX
 OS *Neisseria gonorrhoeae* ATCC 33084.
 XX
 PN MO9640971-A1.
 XX
 PD 19-DEC-1996.
 XX
 PF 03-JUN-1996; 96WO-US08323.
 XX
 PR 07-JUN-1995; 95US-0478140.
 XX
 PA (NEOS-) NEOSE TECHNOLOGIES INC.
 XX
 PI Buczala SL, Johnson KP, Roth S;
 DR WPI; 1997-052351/05.
 DR N-PSDB; AAT49230.
 XX
 PT Transfer of at least 2 saccharide units using

PT poly(glycosyl)transferase - isolated from *N. gonorrhoeae*, catalyses
 PT the addition of both GlcNAc and GalNAc disaccharide(s) units to a
 PT single galactose moiety
 PS Disclosure: Fig 2F-H; 38bp; English.
 XX
 CC A lipo-oligosaccharide-encoding gene region (AAT49230) of *Neisseria*
 CC *gonorrhoeae* ATCC 33084 includes coding sequences for 5 proteins
 CC (AAM06576-80), one of which (AAM06576) is a polyglycosyltransferase
 CC that catalyses the addition of GlcNAc and GalNAc disaccharides to
 CC a galactose moiety. The function of the other 4 proteins is not
 CC stated in the specification.
 CC
 SQ Sequence 280 AA:
 Query Match 99.0%; Score 1425; DB 18; Length 280;
 Best Local Similarity 99.3%; Pred. No. 2.7e-146;
 Matches 278; Conservative 0; Mismatches 2; Indels 0; Gaps 0;
 QY 1 MONHYSLSAERRAHIAADTFGSRGIPFOFPDAMPSERLQAMAEIVPGISAHPYLSG 60
 DB 1 MONHYSLSAERRAHIAADTFGSRGIPFOFPDAMPSERLQAMAEIVPGISAHPYLSG 60
 QY 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLGGAQOFLAEDTWLEERFDKDSAFIVR 120
 DB 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLGGAQOFLAEDTWLEERFDKDSAFIVR 120
 QY 121 LETMFAKYIVRPDKYLVNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPPERIKA 180
 DB 121 LETMFAKYIVRPDKYLVNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPPERIKA 180
 QY 181 VDIAMFTFFDEKGMPIYOVSPALCTQELHYAKFLSONSMGSDLEKDEGRGRRRSLSK 240
 DB 181 VDIAMFTFFDEKGMPIYOVSPALCTQELHYAKFLSONSMGSDLEKDEGRGRRRSLSK 240
 QY 241 VFEDLKRALGKRGREKKRMRQROAELEKYGRVYILFK 280
 DB 241 VFEDLKRALGKRGREKKRMRQROAELEKYGRVYILFK 280
 RESULT 3
 AAR91312
 ID AAR91312 standard; Protein; 279 AA.
 AC AAR91312;
 XX
 DT 09-JUL-1996 (first entry)
 DE *N. gonorrhoeae* glycosyltransferase LgtB.
 XX
 KW Glycosyltransferase; lipo-oligosaccharide; lgt gene; LOS locus;
 KW vaccine.
 XX
 OS *Neisseria gonorrhoeae* strain F62.
 XX
 PN WO9610086-A1.
 XX
 PD 04-APR-1996.
 XX
 PF 25-SEP-1995; 95WO-US12317.
 XX
 PR 26-SEP-1994; 94US-0312387.
 XX
 PA (UYRO) UNIV ROCKEFELLER.
 XX
 PI Gotschlich EC;
 DR WPI; 1996-200924/20.
 DR P-PSDB; AAT14061.
 XX
 PT Nucleic acids encoding glycosyl transferase(s) - used in the
 PT diagnosis of infection with *Neisseria* and for the biosynthesis of
 PT oligo:saccharide(s)

XX Claim 9; Fig 2c; 81pp; English.
 PS 5 Glycosyltransferases (AAR91311-15) are products of the lgt locus
 CC (AAR914061) of *Neisseria gonorrhoeae* strain F62. Glycosyltransferase
 CC LgtB (AAR91312) can be obt'd. by expression of the lgtB coding
 CC sequence in recombinant host cells. A method for adding Gal
 CC beta1-4 to GlcNAc or Glc comprises contacting a reaction mixture
 CC contg. activated Gal to an acceptor moiety comprising a GlcNAc
 CC or Glc residue in the presence of LgtB. Oligosaccharides can be
 CC produced that, when attached to non-toxic lipids, are useful for
 CC *Neisseria* vaccine prepn. Blood group core oligosaccharides, and
 CC mimics of lacto-N-neotetraase, gangliosides and saccharide
 CC portions of glycolipids can also be produced using the enzymes.
 CC
 SQ Sequence 279 AA;
 Query Match 72.0%; Score 1036.5; DB 17; Length 279;
 Best Local Similarity 71.6%; Pred. No. 4.7e-104;
 Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;
 QY 1 MONHYSILASAEERRAHADTFGRGIPPOFPDAMPSERLQAMAEIVPGLSAHPYLSG 60
 DB 1 MONHYSILASAEERRAHADTFGRGIPPOFPDAMPSERLQAMAEIVPGLSAHPYLSG 60
 QY 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLLEGAEQFLAEDTWLEERFDKDSAFIVR 120
 DB 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLLEGAEQFLAEDTWLEERFDKDSAFIVR 120
 QY 121 LETMFAKYIVRPDKVLYNTNRSFPLLESEHCGTAGIISREAMPFLDFAVLPPERIKR 180
 DB 121 LETMFAKYIVRPDKVLYNTNRSFPLLESEHCGTAGIISREAMPFLDFAVLPPERIKR 180
 QY 181 VDLMEFTYFDEKGPVYOVSPALCTOELHYAKFLSONSMLGSDLEKDEQGR----- 233
 DB 181 VDLMEFTYFDEKGPVYOVSPALCTOELHYAKFLSONSMLGSDLEKDEQGR----- 233
 QY 234 -----RHRSILKVMFDLKRALKFGREKKRMREROALERKVGRRVILFK 280
 DB 241 PANTFKHRLRALTKISREKRRQR-----EDLIGKIIVPFQ 279
 RESULT 4
 ID AAY22156 standard; Protein; 275 AA.
 AC AAY22156;
 DT 08-SEP-1999 (first entry)
 DE N. meningitidis Beta-1,4-galactosyltransferase.
 XX
 KW Beta-1,4-galactosyltransferase; lgtB; fusion protein; catalytic domain;
 KM glycosyltransferase; accessory enzyme; nucleotide sugar formation;
 KM saccharide donor; oligosaccharide synthesis;
 KM carbohydrate structure development.
 XX
 OS *Neisseria meningitidis*.
 XX
 PN WO9931224-A2.
 PD 24-JUN-1999.
 XX
 PF 15-DEC-1998; 98WO-CA01180.
 XX
 PR 14-DEC-1998; 98US-0211691.
 PR 15-DEC-1997; 97US-0069443.
 XX
 PA (CANADA) NAT RES COUNCIL CANADA.
 XX
 PI Gilbert M, Makarchuk WW, Young NM;
 XX
 DR WPI; 1999-395174/33.

DR N-PSDB; AAX84281.
 XX
 PT A new glycosyltransferase fusion protein useful in the enzymatic
 PT synthesis of oligosaccharides
 PS Example 2; Fig 2; 63pp; English.
 XX
 CC This sequence represents the *Neisseria meningitidis* Beta-1,4-
 CC galactosyltransferase (also referred to as lgtB). The invention relates
 CC to a nucleic acid encoding a fusion protein that comprises a
 CC glycosyltransferase catalytic domain and a catalytic domain from an
 CC accessory enzyme that is involved in formation of a nucleotide sugar
 CC which is a saccharide donor for a glycosyltransferase reaction. The
 CC fusion protein is useful in the enzymatic synthesis of oligosaccharides.
 CC The fusion proteins are able to catalyze more than one reaction involved
 CC in the enzymatic synthesis. This is useful for the development of
 CC therapeutic agents that have specific carbohydrate structures.
 CC Carbohydrates are involved in recognition elements on the surface of
 CC cells. The fusion protein can be used for the synthesis of both natural
 CC carbohydrates and synthetic derivatives with novel properties. The fusion
 CC polypeptide allows two glycosyltransferase reactions in a single vessel,
 CC provides improved yields of end products. Additionally, cleanup and
 CC disposal of extra solvents and by-products is reduced. The fusion protein
 CC can also use directly different donor analogues and various acceptors
 CC with a terminal galactose residue.
 CC
 SQ Sequence 275 AA;
 Query Match 69.9%; Score 1007; DB 20; Length 275;
 Best Local Similarity 74.7%; Pred. No. 7.5e-101;
 Matches 201; Conservative 20; Mismatches 44; Indels 4; Gaps 1;
 QY 1 MONHYSILASAEERRAHADTFGRGIPPOFPDAMPSERLQAMAEIVPGLSAHPYLSG 60
 DB 1 MONHYSILASAEERRAHADTFGRGIPPOFPDAMPSERLQAMAEIVPGLSAHPYLSG 60
 QY 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLLEGAEQFLAEDTWLEERFDKDSAFIVR 120
 DB 61 VEKACFMSHAVLMEQALDEGLPIYIAVFEDDVLLEGAEQFLAEDTWLEERFDKDSAFIVR 120
 QY 121 LETMFAKYIVRPDKVLYNTNRSFPLLESEHCGTAGIISREAMPFLDFAVLPPERIKR 180
 DB 121 LETMFAKYIVRPDKVLYNTNRSFPLLESEHCGTAGIISREAMPFLDFAVLPPERIKR 180
 QY 181 VDLMEFTYFDEKGPVYOVSPALCTOELHYAKFLSONSMLGSDLEKDEQGR----- 236
 DB 181 VDLMEFTYFDEKGPVYOVSPALCTOELHYAKFLSONSMLGSDLEKDEQGR----- 236
 QY 237 RSLKYVFDLKRALKGRGRRKRRERORQ 265
 DB 241 PANTFKHRLRALTKISREKRRQR-----EDLIGKIIVPFQ 269
 RESULT 5
 ID AAW89331 standard; peptide; 50 AA.
 AC AAW89331;
 DT 26-FEB-1999 (first entry)
 DE *Neisseria meningitidis* lgtE C-terminal peptide.
 XX
 KW *Neisseria meningitidis*; lgtE; lgtB; beta-1,4-galactosyltransferase;
 KM glycosyltransferase; proteolytic enzyme.
 XX
 OS *Neisseria meningitidis*.
 XX
 PN WO9854331-A2.
 PD 03-DEC-1998.
 XX
 PA 26-MAY-1998; 98WO-IB00975.
 XX
 PF

```

XX 27-MAY-1997; 97US-0047751.
PR
XX
PA (CANA ) NAT RES COUNCIL CANADA.
XX
PI Wakarchuk WW, Young NM;
XX
DR WPI: 1999-035177/03.
XX
PS
XX
XX Example 1, Fig 8, 61pp, English.
XX
XX A method has been developed of expressing a glycosyltransferase in a
CC host cell. The method comprises introducing into the host cell a nucleic
CC acid encoding the glycosyltransferase and incubating the host cell under
CC conditions appropriate for expression of the glycosyltransferase, where
CC the host cell substantially lacks a protease that cleaves polypeptides
CC between two consecutively positively charged amino acid residues. The
CC glycosyltransferase can be used in vitro production of
CC oligosaccharide structures which are potential therapeutic agents for
CC use in the manipulation of cell-cell recognition events, particularly
CC adhesion of bacteria and viruses to mammalian cells and leukocyte-
CC endothelial cell interaction through selectins in inflammation. The
CC method provides more readily recoverable active glycosyltransferases
CC than prior art methods involving mammalian glycosyltransferases. The
CC present sequence represents a C-terminal peptide from Neisseria
CC meningitidis 198E from the present invention.
XX
XX Sequence 50 AA:
XX
XX Query Match 16.2%; Score 233; DB 20; Length 50;
XX Best Local Similarity 94.0%; Pred. No. 9,4e-18;
XX Matches 47; Conservative 0; Mismatches 3; Indels 0; Gaps 0;
XX
OY 231 QGRHRRSLKYWFEDIKRALGKFGREKKRMEQRQAELEKYGYRRVILEK 280
DB 1 QERRHRSLSIKWFFDLKRALGKGFRKKRKRMEROARQALEKAGYRRVISPK 50
XX
XX RESULT 6
XX ABB62795
XX ID ABB62795 standard; Protein; 1128 AA.
XX AC ABB62795;
XX AD
XX DT 26-MAR-2002 (first entry)
XX DE Drosophila melanogaster polypeptide SEQ ID NO 15177.
XX KW Drosophila; developmental biology; cell signalling; insecticide;
XX pharmaceutical.
XX OS Drosophila melanogaster.
XX PN WO200171042-A2.
XX PD
XX PE 27-SEP-2001.
XX PF 23-MAR-2001; 2001WO-US09231.
XX PR 23-MAR-2000; 2000US-191637P.
XX PR 11-JUL-2000; 2000US-0614150.
XX PA (PEKE ) PE CORP NY.
XX PI Venter JC, Adams M, Li PWD, Myers EW;
XX WPI: 2001-656860/75.
XX DR N-PSDB; ABLO6898.
XX

```

[illegible]

PR 19-MAY-2000; 2000US-205515P.
 XX (HUMA-) HUMAN GENOME SCI INC.
 XX
 XX Birse CE, Rosen CA;
 DR WPI: 2002-122018/16.
 DR N-PSDB; ABL89862.
 XX
 PT Novel 1405 isolated polypeptides, useful for diagnosis, treatment and
 PT prevention of neural, immune system, muscular, reproductive,
 PT gastrointestinal, pulmonary, cardiovascular, renal and proliferative
 PT disorders -
 PS
 PS Claim 11; SEQ ID NO 1829; 2081pp + Sequence Listing; English.
 XX
 CC The invention relates to novel genes (ABL89449-ABL90833) and proteins
 CC (ABB89040-ABB90444) useful for preventing, treating or ameliorating
 CC medical conditions e.g. by protein or gene therapy. The genes are
 CC isolated from a range of human tissues disclosed in the specification.
 CC The nucleic acids, proteins, antibodies and (ant)agonists are useful
 CC in the diagnosis, treatment and prevention of: (a) cancer, e.g. breast
 CC and ovarian cancer and other cancers of the adrenal gland, bone, bone
 CC marrow, breast, gastrointestinal tract, liver, lung, or urogenital;
 CC (b) immune disorders e.g. Addison's disease, allergies, autoimmune
 CC haemolytic anaemia, autoimmune thyroiditis, diabetes mellitus, Crohn's
 CC disease, multiple sclerosis, rheumatoid arthritis and ulcerative
 CC colitis; (c) cardiovascular disorders such as myocardial ischaemia;
 CC (d) wound healing; (e) neurological diseases e.g. cerebral anoxia and
 CC epilepsy; and (f) infectious diseases such as viral, bacterial, fungal
 CC and parasitic infections.
 CC Note: The sequence data for this patent did not form part of the
 CC printed specification, but was obtained in electronic format directly
 CC from WIPO at ftp.wipo.int/pub/published_pct_sequences.
 CC
 XX
 XX Sequence 423 AA:
 XX
 XX Query Match 9.0%; Score 129.5; DB 23; Length 423;
 XX Best Local Similarity 25.1%; Pred. No. 3.6e-05;
 XX Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.
 OS
 OS Homo sapiens.
 XX
 XX WO200058473-A2.
 XX
 PD 05-OCT-2000.
 XX
 PD 31-MAR-2000; 2000WO-0508621.
 XX
 PF 31-MAR-1999; 99US-0127607.
 PR 02-APR-1999; 99US-0127636.
 PR 05-APR-1999; 99US-0127728.
 PR 30-MAR-2000; 2000US-0540763.
 XX
 PA (CURA-) CURAGEN CORP.
 XX
 PI Shinkets RA, Leach M;
 DR WPI: 2000-602362/57.
 DR N-PSDB; AAC73999.
 XX
 PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 XX
 PS Claim 11; Page 2331-2332; 5507pp; English.
 XX
 CC AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnerary;
 CC antiproliferic; antiparkinsonian; nootropic; neuroprotective;
 CC osteopathic; anticonvulsant; antiallergic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.
 CC
 XX
 XX Sequence 517 AA:
 XX
 XX Query Match 9.0%; Score 129.5; DB 21; Length 517;
 XX Best Local Similarity 25.1%; Pred. No. 4.8e-05;
 XX Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

Db 364 DLIIYLRKQ-----VNPEKETAVEGLPGLVAVAGSYWTLAVALRLAGARKILLASQPLR 416
QY 165 FFL--DREAVL-----PPERIKAVDLMMFTYFFDKMGMPYQVSPALCTQELHYA 212
Db 417 RMLPVDEFLPIMFDQHPNEQYKA-----HEWPRD-LVAFSAQPLLA-PTHYA 462

RESULT 9
AAM93202

ID AAM93202 standard; Protein; 517 AA.
AC AAM93202;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 2591.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-524255/58.
DR N-PSDB; AAK94110.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 2591; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 517 AA;

Query Match 9.0%; Score 129.5; DB 22; Length 517;
Best Local Similarity 25.1%; Pred. No. 4.8e-05;
Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

QY 5 VISLAAERRAHIADTFGSRGIPFOFPA-----LMPSERLBOAMAEIVPGLSAHRY--- 57
Db 246 VISLARRPDRRERMLASLWMEISGRVVDVADGWMNSSAIRNLGYDLPLPGYO-DPYSGR 304
QY 58 -LSGVKACFMSHAVLMEQALDEGLPIYAVFEDDVLG---EGAEQFLAEDTWLEERFDK 113
Db 305 TLTGKEVGCTLSHYSTWEEVYARGLARVLVFEEDVAFESNFRRLRLMEDV-EAEKLSW 363
QY 114 DSAFIVRLTETMAKVIVRPDKVLNENRSFPLESEHCGTAGY-----IISREAMR 164

Db 364 DLIIYLRKQ-----VNPEKETAVEGLPGLVAVAGSYWTLAVALRLAGARKILLASQPLR 416
QY 165 FFL--DREAVL-----PPERIKAVDLMMFTYFFDKMGMPYQVSPALCTQELHYA 212
Db 417 RMLPVDEFLPIMFDQHPNEQYKA-----HEWPRD-LVAFSAQPLLA-PTHYA 462

RESULT 10
AAM93206

ID AAM93206 standard; Protein; 517 AA.
AC AAM93206;
XX
DT 06-NOV-2001 (first entry)
XX
DE Human polypeptide, SEQ ID NO: 2599.
XX
KW Human; full length cDNA; cDNA synthesis; oligo-capping.
XX
OS Homo sapiens.
XX
PN EP130094-A2.
XX
PD 05-SEP-2001.
XX
PF 07-JUL-2000; 2000EP-0114089.
XX
PR 08-JUL-1999; 99JP-0194486.
PR 11-JAN-2000; 2000JP-0118774.
PR 02-MAY-2000; 2000JP-0183765.
XX
PA (HELI-) HELIX RES INST.
XX
PI Ota T, Nishikawa T, Isogai T, Hayashi K, Ishii S, Kawai Y;
PI Wakamatsu A, Sugiyama T, Nagai K, Kojima S, Otsuki T, Koga H;
XX
DR WPI: 2001-524255/58.
DR N-PSDB; AAK94114.
XX
PT 830 Primers useful for synthesizing full length cDNA clones and their
PT use in genetic manipulation -
XX
PS Claim 8; SEQ ID NO 2599; 1380bp + sequence listing; English.
XX
CC The invention relates to primers for synthesizing full length cDNA
CC clones. 830 cDNA molecules encoding a human protein have been
CC isolated and nucleotide sequences of 5'- and 3'-ends of the cDNA
CC molecules have been determined. Primers for synthesizing the full length
CC cDNA are useful for clarifying the function of the protein encoded by
CC the cDNA. The full length clones were obtained by construction of full
CC length enriched cDNA libraries that were synthesised by the oligo-capping
CC method. The primers enable the production of the full length cDNA easily
CC without any special methods. The present sequence is a polypeptide
CC encoded by a full length human cDNA of the invention.
CC Note: The sequence data for this patent did not form part of the printed
CC specification, but was obtained in CD-ROM format directly from EPO.
XX
SQ Sequence 517 AA;

Query Match 9.0%; Score 129.5; DB 22; Length 517;
Best Local Similarity 25.1%; Pred. No. 4.8e-05;
Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

QY 5 VISLAAERRAHIADTFGSRGIPFOFPA-----LMPSERLBOAMAEIVPGLSAHRY--- 57
Db 246 VISLARRPDRRERMLASLWMEISGRVVDVADGWMNSSAIRNLGYDLPLPGYO-DPYSGR 304
QY 58 -LSGVKACFMSHAVLMEQALDEGLPIYAVFEDDVLG---EGAEQFLAEDTWLEERFDK 113
Db 305 TLTGKEVGCTLSHYSTWEEVYARGLARVLVFEEDVAFESNFRRLRLMEDV-EAEKLSW 363
QY 114 DSAFIVRLTETMAKVIVRPDKVLNENRSFPLESEHCGTAGY-----IISREAMR 164

Db 364 DLIIYGRQ-----VPEKETAVEGLPGILVAGSYWTLAYALRLAGARKLLASQPLR 416
 QY 165 FFL--DREAVL-----PPERIKAVDLMFTYFDEKGMPPVYVSPALCTQELHYA 212
 Db 417 RMLPVDEFLLPMEQHPNDQYKA-----HFWRPD-LVAFSAQPLLAA-PTHYA 462

RESULT 11

AAG81257
 ID AAG81257 standard; Protein: 578 AA.

AC AAG81257;

DT 10-SEP-2001 (first entry)

XX Human AFP protein sequence SEQ ID NO:32.

DE Human AFP protein sequence SEQ ID NO:32.

XX Human; secreted protein; secretion; bacterial cell; fungal cell;

KW eukaryotic cell; fusion protein; maltose binding protein;

KM immunoglobulin constant region; polyhistidine tag.

XX Homo sapiens.

PN WO200129221-A2.

PD 26-APR-2001.

PE 20-OCT-2000; 2000WO-US29052.

PR 20-OCT-1999; 99US-0160712.

XX (ZYMO) ZYMOGENETICS INC.

PI Conklin DC, Yee DP;

DR WPI: 2001-300340/31.

DR N-PSDB; AAH52108.

XX Isolated polypeptide for directing secretion of proteins of interest

PT from a host cell including, e.g. bacteria, includes contiguous amino

PT acid residues of polypeptide with specified amino acids

CC Claim 1; Page 121-122; 617pp; English.

CC AAH52093 to AAH52303 encode the human secreted proteins given in AAG81242

CC to AAG81453. The secreted proteins can be used for directing the

CC secretion of proteins of interest from a host cell including bacteria,

CC fungal cells, and cultured higher eukaryotic cells. The present invention

CC also describes fusion proteins, where a secreted protein of the invention

CC is operably linked via a peptide bond or peptide linker to a second

CC protein selected from the group consisting of maltose binding protein,

CC an immunoglobulin constant region, a polyhistidine tag and a peptide

CC given in AAG81453.

CC Sequence 578 AA;

SO Query Match 9.0%; Score 129.5; DB 22; Length 578;

Best Local Similarity 25.1%; Pred. No. 5.6e-05;

Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

QY 165 FFL--DREAVL-----PPERIKAVDLMFTYFDEKGMPPVYVSPALCTQELHYA 212
 Db 495 RMLPVDEFLLPMEQHPNDQYKA-----HFWRPD-LVAFSAQPLLAA-PTHYA 540

RESULT 12

AAV30812
 ID AAV30812 standard; Protein: 579 AA.

AC AAV30812;

DT 12-OCT-1999 (first entry)

XX Human secreted protein encoded from gene 2.

DE Secreted protein; prevention; treatment; protein therapy; gene therapy;

KW diagnosis; cancer; tumour; neurodegenerative disorder; blood disorder;

KM developmental abnormality; fetal deficiency; leukemia; autoimmune; acne;

KW hepatic disease; renal disease; lymphoma; inflammation; allergy; asthma;

KM Alzheimer's disease; cognitive disorder; schizophrenia; obesity; sepsis;

KW osteoporosis; arthritis; infection; AIDS; connective tissue disorder;

KM transplant rejection; diabetes; psoriasis; cardiovascular disorder;

KM reproductive disorder; food additive; food preservative; human; primer;

KM early promoter; GAS; gamma activation element.

XX Homo sapiens.

PN WO9940100-A1.

PD 12-AUG-1999.

PE 04-FEB-1999; 99WO-US02293.

PR 09-FEB-1998; 98US-0074341.

PR 09-FEB-1998; 98US-0074037.

PR 09-FEB-1998; 98US-0074118.

PR 09-FEB-1998; 98US-0074141.

PR 09-FEB-1998; 98US-0074157.

XX (HOMA-) HUMAN GENOME SCI INC.

PI Kyaw H, Lafleur DW, Moore PA, Rosen CA, Ruben SM;

PI Shi Y, Wei Y;

DR WPI: 1999-479426/40.

DR N-PSDB; AAZ00803.

XX New isolated human genes potentially useful for, e.g. developmental

PT abnormalities and fetal deficiencies

CC Claim 1b; Page 206-208; 263pp; English.

CC This invention describes novel isolated human genes and the secreted

CC proteins they encode. The polynucleotides and their corresponding

CC secreted polypeptides are useful for preventing, treating or ameliorating

CC medical conditions e.g. by protein or gene therapy. Also pathological

CC conditions can be diagnosed by determining the amount of the new

CC polypeptides in a sample or by determining the presence of mutations in

CC the new polynucleotides. Specific uses are described for the

CC polynucleotides of the invention based on which tissues they are most

CC highly expressed in, and include developing products for the diagnosis or

CC treatment of cancer, tumours, neurodegenerative disorders, developmental

CC abnormalities and fetal deficiencies, blood disorders, leukemias,

CC diseases of the immune system, autoimmune diseases, hepatic and renal

CC disease, lymphomas, inflammation, allergies, Alzheimer's and cognitive

CC disorders, schizophrenia, obesity, osteoporosis, arthritis, infections,

CC AIDS, connective tissue disorders, transplant rejection, diabetes,

CC asthma, sepsis, acne, psoriasis, cardiovascular disorders, and

CC reproductive disorders. The polypeptides or polynucleotides can also be

CC used as food additives or preservatives. The polypeptide are also useful

CC for identifying their binding partners. This sequence represents a

CC secreted protein described in the invention.

```

SQ      Sequence      579 AA:
Query Match      9.0%; Score 129.5; DB 20; Length 579;
Best Local Similarity 25.1%; Pred. No. 5.6e-05;
Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

OY      5 VISLASAERRAHIAIDFTSSRGIPFOFPDA---LMPSERLDAQMAELVPGLSAHRY--- 57
        ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      324 VISLARPRRERREMLASLWEMETISGRVDAVDGMLNNSAIRNLGVDLLPGYQ-DPYSGR 382
OY      58 -LSGVKACFMASHAVLMEQALDEGLPIYAVFEDVDLLG---EGAEQFLAEDPTLLEERFDK 113
        : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      383 TLTKEVGCGFLSHYSIMEEVARGLARVLVEFDDVRFESNFRGRLERLMEDV-EAEKLSW 441
OY      114 DSAFIYRLFTMFPAKVIYRPDKVLYENRSFPLLESEHCCTAGY-----IISREAMR 164
DB      442 DLIYLGKQ-----VNPEKETAVEGLPGLVAGYSYWTLYALRLAGARKLLASQPLR 494
OY      165 FFL--DRAVL-----PPERIKAVDLMFTYFPDKGMPYQVSPALCQELHYA 212
DB      495 RMLPVDEFPLPMDQHPNEQYKA-----HEWPRD-LVAFSAQPLAA-PTHYA 540

RESULT 13
AAB12121
ID      AAB12121 standard; Protein; 595 AA.
XX
AC      AAB12121;
XX
DT      02-FEB-2001 (first entry)
XX
DE      Hydrophobic domain protein from clone HP02962 isolated from KB cells.
XX
KW      Human; secreted protein; membrane protein; hydrophobic domain;
KW      proliferation control; differentiation induction; material transport;
KW      biophylaxis; signal receptor; ion channel; transporter; immunostimulant;
KW      immunosuppressant; haematopoiesis regulator; chemotactic; chemokine;
KW      haemolytic; thrombolytic; anti-inflammatory; tumour inhibition;
KW      autoimmune disease; Alzheimer's disease; Parkinson's disease; cancer.
XX
OS      Homo sapiens.
XX
PN      WO200029448-A2.
XX
PD      25-MAY-2000.
XX
PE      17-NOV-1999; 99WO-JP06412.
XX
PF      17-NOV-1998; 98UP-0326255.
XX
PR      22-DEC-1998; 98UP-0364315.
XX
PR      16-MAR-1999; 99UP-0069811.
XX
PR      27-APR-1999; 99UP-0119299.
XX
PR      19-MAY-1999; 99UP-0138169.
XX
PA      (SAGA ) SAGAMI CHEM RES CENT.
XX
PA      (PROT-) PROTEGENE INC.
XX
PI      Kato S, Kimura T;
XX
PI      WPI: 2000-387753/73.
XX
DR      N-PSDB; AAA60183, AAA60193.
XX
PT      Proteins comprising hydrophobic regions, such as secretory and membrane
PT      proteins, useful in research and diagnostics and having various
PT      activities e.g. immunomodulatory, anti-inflammatory, chemokine,
PT      hemostatic, thrombolytic -
XX
PS      Claim 1; Page 184-186; 410pp; English.
XX
CC      Secretory proteins play important roles in the proliferation control, the
CC      differentiation induction, the material transport and the biophylaxis of
CC      cells. Membrane proteins have important roles as signal receptors, ion
CC      channels and transporters. The present sequence is a human protein which
```

```

CC      has at least one hydrophobic domain. This protein may be a secretory or a
CC      membrane protein. The present protein may have cytokine and cell
CC      proliferation/differentiation activity, immune stimulating or suppressing
CC      activity, haematopoiesis activity, tissue growth activity,
CC      activin/inhibin activity, chemotactic/chemokine activity, haemostatic
CC      and thrombolytic activity, anti-inflammatory activity and tumour
CC      inhibition activity. The present protein could therefore be used for
CC      treatment of autoimmune disease, Alzheimer's disease, Parkinson's
CC      disease, and cancer.
XX
SQ      Sequence      595 AA:
Query Match      9.0%; Score 129.5; DB 21; Length 595;
Best Local Similarity 25.1%; Pred. No. 5.8e-05;
Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

OY      5 VISLASAERRAHIAIDFTSSRGIPFOFPDA---LMPSERLDAQMAELVPGLSAHRY--- 57
        ||||| : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      324 VISLARPRRERREMLASLWEMETISGRVDAVDGMLNNSAIRNLGVDLLPGYQ-DPYSGR 382
OY      58 -LSGVKACFMASHAVLMEQALDEGLPIYAVFEDVDLLG---EGAEQFLAEDPTLLEERFDK 113
        : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB      383 TLTKEVGCGFLSHYSIMEEVARGLARVLVEFDDVRFESNFRGRLERLMEDV-EAEKLSW 441
OY      114 DSAFIYRLFTMFPAKVIYRPDKVLYENRSFPLLESEHCCTAGY-----IISREAMR 164
DB      442 DLIYLGKQ-----VNPEKETAVEGLPGLVAGYSYWTLYALRLAGARKLLASQPLR 494
OY      165 FFL--DRAVL-----PPERIKAVDLMFTYFPDKGMPYQVSPALCQELHYA 212
DB      495 RMLPVDEFPLPMDQHPNEQYKA-----HEWPRD-LVAFSAQPLAA-PTHYA 540

RESULT 14
ABP42016
ID      ABP42016 standard; Protein; 636 AA.
XX
AC      ABP42016;
XX
DT      22-AUG-2002 (first entry)
XX
DE      Human ovarian antigen HERKF51, SEQ ID NO:3148.
XX
KW      Human; ovarian antigen; ovary; ovarian; breast; cancer; tumour;
KW      ovarian cancer; breast cancer; tumour; reproductive system disorder;
KW      infertility; pregnancy disorder; anovulation; polycystic ovary syndrome;
KW      PCOS; ovarian cyst; dysmenorrhoea; endocrine disorder; infection;
KW      inflammatory condition; immune disorder; blood disorder;
KW      cardiovascular disorder; respiratory disorder; neurological disorder;
KW      gastrointestinal disorder; urinary system disorder; drug screening;
KW      gene therapy; chromosome mapping; forensic analysis;
KW      antibody preparation; cytostatic; immunomodulatory; neuroprotective;
KW      antiinflammatory; gynaecological; reproductive.
XX
OS      Homo sapiens.
XX
PN      WO200200677-A1.
XX
PD      03-JAN-2002.
XX
PE      07-JUN-2001; 2001WO-US18569.
XX
PR      07-JUN-2000; 2000US-209467P.
XX
PA      (HUMA-) HUMAN GENOME SCT INC.
XX
PI      Birse CE, Rosen CA;
XX
PI      WPI: 2002-147878/79.
XX
DR      N-PSDB; AB055093.
XX
PT      Isolated nucleic acid molecules encoding novel ovarian polypeptides,
PT      useful in the prevention, treatment and diagnosis of cancer (e.g.
```

PT ovarian cancer), immune disorders, cardiovascular disorders and
 PT neurological diseases -
 XX
 PS Claim 11: SEQ ID NO 3148; 2922pp; English.
 XX
 CC The invention relates to 2175 novel human ovarian antigens (ABP41054-
 CC ABP43228) and to cDNAs encoding them (AB054131-AB056305), and also
 CC encompasses polypeptides 90% identical and polynucleotides 95% identical
 CC to the sequences of the invention. The invention additionally relates to
 CC recombinant vectors and host cells comprising human ovarian antigen
 CC polynucleotides, antibodies against human ovarian antigens, and the use
 CC of ovarian antigen polynucleotides and polypeptides in diagnosing,
 CC treating, prognosing or preventing various ovarian and/or breast-related
 CC disorders. Such conditions include ovarian cancer and breast cancer, and
 CC metastatic tumours of ovarian or breast origin, reproductive system
 CC disorders (e.g., infertility, disorders of pregnancy, anovulation,
 CC polycystic ovary syndrome, ovarian cysts, and dysmenorrhea), endocrine
 CC disorders, infections (e.g., chlamydia, HIV, toxoplasmosis), and toxic
 CC shock syndrome), inflammatory conditions (e.g., mastitis, oophoritis and
 CC vaginitis), immune disorders (e.g., congenital and acquired
 CC immunodeficiencies, autoimmune oophoritis, systemic lupus erythematosus),
 CC blood-related disorders (e.g., anaemia), cardiovascular disorders,
 CC respiratory disorders, neurological disorders, gastrointestinal disorders
 CC and urinary system disorders. Ovarian antigen polypeptides and
 CC polynucleotides may also be used in screening for compounds which
 CC modulate ovarian antigen expression or activity. The polynucleotides may
 CC further be used for gene therapy, chromosome mapping, in the
 CC identification of individuals and in forensic analysis, and the
 CC polypeptides may be used as food additives or to prepare antibodies
 CC useful in disease diagnosis, drug targeting and phenotyping. The present
 CC sequence represents a human ovarian antigen of the invention.
 CC Note: The sequence data for this patent did not form part of the printed
 CC specification, but was obtained in electronic format directly from WIPO
 CC at [ftp.wipo.int/pub/published_sequences](http://wipo.int/pub/published_sequences).

XX Sequence 636 AA;

Query Match 9.0%; Score 129.5; DB 23; Length 636;
 Best Local Similarity 25.1%; Pred. No. 6.4e-05;
 Matches 59; Conservative 36; Mismatches 95; Indels 45; Gaps 12;

OY 5 VISLASAERRAHNADTFGSGCIPQPFDA---LMPSERLQMAELVPLSLAHPY--- 57
 DB 365 VISLARPRDREREMRLASLMEWISGRVVDVADGWMNLSAIRNLGVDDLPGYQ-DPYSGR 423
 OY 58 -LSGVERAKCEHSHAVMEQALDEGLPIYAVFEDDYILG---EGAEQFLAEPTWLEERDK 113
 DB 424 TLTIQGEVGFSLSHSIMEEVAARGLARLVPEDDVRESNFRGLERLMDV-EAEKLSW 482
 OY 114 DSAFIVRLTFMFAVIYAPDKVLNENRSPFLSEHC GTAGY-----ITSREAMR 164
 DB 483 DLITIGRKQ-----VNPKEKTAVEGLPIGLVAVGYWTIATALRLAGAKKLASQPLR 535
 OY 165 FFL--DREPAVL-----PERIKAVDLMMFTYFDEKGPYQVSPALCTQELHYA 212
 DB 536 RMLPVDFELPIMFQHNPEQYKA-----HFMPRD-LVAFSAQPLLA-PTHYA 581

RESULT 15

AAB41900 AAB41900 standard; Protein: 521 AA.

XX AAB41900;

XX 08-FEB-2001 (first entry)

DE Human ORFX ORF1664 polypeptide sequence SEQ ID NO:3328.

XX Human, open reading frame; ORFX: detection; cytostatic; hepatotropic;
 KW vulnery; antiposrotatic; antiparkinsonian; neurotropic; neuroprotective;
 KW anticonvulsant; osteopethic; antiathrilitic; immunosuppressant; cardiant;
 KW immunostimulant; thrombolytic; coagulant; vasotropic; antiadabetic;
 KW hypotensive; dermatological; immunosuppressive; antiinflammatory;

KW antiviral; antibacterial; antifungal; antirheumatic; antithyroid;
 KW antianemic; gene therapy; cancer; proliferative disorder; hypertension;
 KW neurodegenerative disorder; osteoarthritis; graft vs host disease;
 KW cardiovascular disease; diabetes mellitus; hypothyroidism; SCID; AIDS;
 KW cholesterol ester storage; systemic lupus erythematosus; infection;
 KW severe combined immunodeficiency; malaria; autoimmune disorder; asthma;
 KW allergy; aplastic anaemia; nocturnal haemoglobinuria; burn; wound;
 KW bone damage; cartilage damage; antiinflammatory disease; coagulation;
 KW thrombosis; contraceptive.

OS Homo sapiens.

PN W0200058473-A2.

PD 05-OCT-2000.

PF 31-MAR-2000; 2000WO-US08621.

PR 31-MAR-1999; 99US-0127607.

PR 02-APR-1999; 99US-0127636.

PR 05-APR-1999; 99US-0127728.

PR 30-MAR-2000; 2000US-0540763.

PA (CURA-) CURAGEN CORP.

PI Shinkets RA, Leach M;

DR WPI: 2000-602362/57.

DR N-PSDB: AAC76109.

PT Novel nucleic acids and peptides derived from open reading frame X,
 PT useful for treating e.g. cancers, proliferative disorders,
 PT neurodegenerative disorders and cardiovascular disease -
 PT
 XX
 PS Claim 11; Page 2519-2520; 5507pp; English.

XX AAC74446 to AAC77606 encode the proteins given in AAB40237 to AAB43397,
 CC which represent the human ORFX open reading frames 1 to 3161. The ORFX
 CC sequences have activities such as: cytostatic; hepatotropic; vulnery;
 CC antiposrotatic; antiparkinsonian; neurotropic; neuroprotective;
 CC osteopethic; anticonvulsant; antiathrilitic; immunosuppressant;
 CC immunostimulant; cardiant; thrombolytic; coagulant; vasotropic;
 CC antidiabetic; hypotensive; dermatological; immunosuppressive;
 CC antiinflammatory; antibacterial; antiviral; antifungal; antirheumatic;
 CC antithyroid; and antianemic. The sequences can be used for determining
 CC the presence of or predisposition to, or preventing or treating
 CC pathological conditions associated with an ORFX-associated disorder. The
 CC nucleic acids can be used to express ORFX proteins in gene therapy
 CC vectors. The proteins and nucleic acids may be used to treat cancers,
 CC proliferative disorders, neurodegenerative disorders, osteoarthritis,
 CC graft vs host disease, cardiovascular disease, diabetes mellitus,
 CC hypertension, hypothyroidism, cholesterol ester storage, systemic lupus
 CC erythematosus, severe combined immunodeficiency (SCID), AIDS, viral,
 CC bacterial or fungal infection, malaria, autoimmune disorders, asthma,
 CC allergies, aplastic anaemia, burns, wounds, bone and cartilage damage,
 CC nocturnal haemoglobinuria, antiinflammatory disease; to enhance
 CC coagulation; to inhibit thrombosis; and as a contraceptive.

XX Sequence 521 AA;

Query Match 8.8%; Score 127; DB 21; Length 521;
 Best Local Similarity 23.8%; Pred. No. 9e-05;
 Matches 69; Conservative 45; Mismatches 120; Indels 56; Gaps 15;

OY 5 VISLASAERRAHNADTFGSGCIPQPFDA---MPSERLQMAELVPLSLAHPY--- 57

DB 244 MINLRRODREREMRLAQAQIEICRLVEADVGRAMNTSQVEALGIQMLPGY-RDPYGR 302

OY 58 -LSGVERAKCEHSHAVMEQALDEGLPIYAVFEDDYILGEGAEQF-----LAETWLEE 109

DB 303 PLTIGELGCLFSLNINIKMEVVDRLQKSLVEEDDRF---EIFFKRLMLNIRV-ERE 357

OY 110 RFDSDAFIVRLTFMFAVIYAPDKVLNENRSPFLSEHC-GTAGYIISREAMRFLD 168

Tue Dec 3 09:23:15 2002

us-10-007-267-6.rag

Page 10

Db	358	GLMDMLIYVGKRMQ----	YENHPKAV----	PRVNNILVEADYISWTTLAVYISLGARKLL-	409
Oy	169	RFAYLPRERIKAVDLMMFTYTFEDKEGMPYOVSPALCTOELH-----	YAKFLSÖNS	219	
Db	410	--AAEPLESKMLPRVEEFL-PRVNEDEH--PRVEEYKAHFSLRMLHAESVPELLIYPRHYTGD	464		
Oy	220	MLGSDLEKDEPQGRHRRSLKLVMEFLKRALGKFGEREKKHMERQAELE	269		
Db	465	GYVSDTEFSVYNNH-----YKTDMD-----	RASQKMRQDALSRE	502	

```
Search completed: December 2, 2002, 11:59:04
Job time : 31.0623 secs
```


GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 : Search time 10.4223 Seconds
(without alignments)
790.458 Million cell updates/sec

Title: US-10-007-267-6
Perfect score: 1440
Sequence: 1 MONHVISLASAERRAHAD.....ERQROAELEKVGGRVILFK 280

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 262574 seqs, 29422922 residues

Total number of hits satisfying chosen parameters: 262574

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :

Issued_Patents_AA:*
1: /cgn2_6/ptodata/1/1aa/5A_COMB.pep:*
2: /cgn2_6/ptodata/1/1aa/5B_COMB.pep:*
3: /cgn2_6/ptodata/1/1aa/6A_COMB.pep:*
4: /cgn2_6/ptodata/1/1aa/6B_COMB.pep:*
5: /cgn2_6/ptodata/1/1aa/PCUTS_COMB.pep:*
6: /cgn2_6/ptodata/1/1aa/Backfilest.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1440	100.0	280	1 US-08-312-387B-6	Sequence 6, Appli
2	1440	100.0	280	1 US-08-683-426-6	Sequence 6, Appli
3	1440	100.0	280	1 US-08-683-458-6	Sequence 6, Appli
4	1440	100.0	280	2 US-08-878-360-6	Sequence 6, Appli
5	1440	100.0	280	3 US-08-478-140B-6	Sequence 6, Appli
6	1440	100.0	280	4 US-09-333-412-6	Sequence 6, Appli
7	1440	100.0	280	4 US-09-338-943-6	Sequence 6, Appli
8	1036.5	72.0	279	1 US-08-312-387B-8	Sequence 8, Appli
9	1036.5	72.0	279	1 US-08-683-426-8	Sequence 8, Appli
10	1036.5	72.0	279	1 US-08-683-458-8	Sequence 8, Appli
11	1036.5	72.0	279	2 US-08-878-360-8	Sequence 8, Appli
12	1036.5	72.0	279	4 US-09-333-412-8	Sequence 8, Appli
13	99	6.9	394	2 US-08-646-580B-40	Sequence 40, Appli
14	99	6.9	394	4 US-09-412-184-40	Sequence 40, Appli
15	83	5.8	712	2 US-08-468-576B-17	Sequence 17, Appli
16	83	5.8	712	2 US-08-468-579B-17	Sequence 17, Appli
17	83	5.8	712	3 US-08-468-577B-17	Sequence 17, Appli
18	81.5	5.7	664	1 US-08-485-284A-5	Sequence 5, Appli
19	79.5	5.5	1098	4 US-08-923-982A-8	Sequence 8, Appli
20	78.5	5.5	1104	4 US-08-923-992A-4	Sequence 4, Appli
21	78	5.4	337	4 US-09-134-001C-3799	Sequence 3799, Ap
22	76.5	5.3	1388	4 US-09-572-191-2	Sequence 2, Appli
23	76.5	5.3	1388	4 US-09-723-262-2	Sequence 2, Appli
24	76.5	5.3	1388	4 US-09-723-262-2	Sequence 2, Appli
25	75.5	5.2	1128	4 US-08-923-992A-6	Sequence 6, Appli
26	75.5	5.2	1227	1 US-08-448-170-8	Sequence 8, Appli
27	75.5	5.2	1227	4 US-08-961-803-9	Sequence 9, Appli

28	75	5.2	337	3 US-09-032-372-2	Sequence 2, Appli
29	75	5.2	1375	4 US-09-722-139-2	Sequence 2, Appli
30	75	5.2	1375	4 US-09-721-832-2	Sequence 2, Appli
31	75	5.2	1375	4 US-09-721-689-2	Sequence 2, Appli
32	74.5	5.2	506	4 US-09-134-001C-4049	Sequence 4049, Ap
33	74	5.1	344	6 5210183-2	Patent No. 5210183
34	74	5.1	630	4 US-09-360-545-67	Sequence 67, Appli
35	74	5.1	683	6 5210183-3	Patent No. 5210183
36	73.5	5.1	487	4 US-09-724-224-8	Sequence 8, Appli
37	73.5	5.1	512	4 US-09-724-224-4	Sequence 4, Appli
38	73.5	5.1	539	3 US-09-057-969-4	Sequence 4, Appli
39	73.5	5.1	624	3 US-09-057-969-3	Sequence 3, Appli
40	73.5	5.1	954	3 US-09-057-969-2	Sequence 3, Appli
41	73.5	5.1	984	1 US-08-242-932-2	Sequence 2, Appli
42	73.5	5.1	984	1 US-08-714-481-2	Sequence 2, Appli
43	73.5	5.1	984	5 PCT-US95-06111-2	Sequence 2, Appli
44	73.5	5.1	1164	4 US-08-923-992A-2	Sequence 2, Appli
45	73	5.1	969	1 US-08-365-689-3	Sequence 3, Appli

ALIGNMENTS

```
RESULT 1
US-08-312-387B-6
; Sequence 6, Application US/08312387B
; Patent No. 5545553
;
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESS: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/312,387B
FILING DATE: July 7, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO.: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-312-387B-6
Query Match 100.0%; Score 1440; DB 1: Length 280;
Best Local Similarity 100.0%; Pred. No. 7e-155;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
```

OY 1 MONHVISLASAERRAHADTFTGSGRIGIPPOFPDAMPSERLQANAEVPGISAHPIYSG 60
DB 1 MONHVISLASAERRAHADTFTGSGRIGIPPOFPDAMPSERLQANAEVPGISAHPIYSG 60
OY 61 VKACFMASHAVLMQALDEGLPYIAVFEDDVLGEGAGQFLAEDTWLBERFDKDSAFIVR 120

Db 61 VERACFMSHAVIWEQALDGLPIYIAVFEDVLLGGAEOFLAEDTWLEERFKDSAFIVR 120
QY 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180
Db 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180
QY 181 VDLMMFTYFDEKGMVYQVSPALCTQELHYAKFLSÖNSMLGSDLEKDEÖGRHRRSLK 240
Db 181 VDLMMFTYFDEKGMVYQVSPALCTQELHYAKFLSÖNSMLGSDLEKDEÖGRHRRSLK 240
QY 241 VMFDLKRALGKFGREKKRMRÖRQAELEKVGGRVILFK 280
Db 241 VMFDLKRALGKFGREKKRMRÖRQAELEKVGGRVILFK 280

RESULT 2

US-08-683-426-6
; Sequence 6, Application US/08683426
; Patent No. 5705367
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauder & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,426
; FILING DATE:
; CLASSIFICATION: 536
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 536
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REFERENCE/DOCKET NUMBER: 26,742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-683-426-6

Query Match 100.0%; Score 1440; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 7e-155;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MONHVISTLSAERRAHINDTFGSRGIPQFPDAMPSERLBOAMAEVPGLSAHPYLSG 60
Db 1 MONHVISTLSAERRAHINDTFGSRGIPQFPDAMPSERLBOAMAEVPGLSAHPYLSG 60
QY 61 VERACFMSHAVIWEQALDGLPIYIAVFEDVLLGGAEOFLAEDTWLEERFKDSAFIVR 120
Db 61 VERACFMSHAVIWEQALDGLPIYIAVFEDVLLGGAEOFLAEDTWLEERFKDSAFIVR 120

QY 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180
Db 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180
QY 181 VDLMMFTYFDEKGMVYQVSPALCTQELHYAKFLSÖNSMLGSDLEKDEÖGRHRRSLK 240
Db 181 VDLMMFTYFDEKGMVYQVSPALCTQELHYAKFLSÖNSMLGSDLEKDEÖGRHRRSLK 240
QY 241 VMFDLKRALGKFGREKKRMRÖRQAELEKVGGRVILFK 280
Db 241 VMFDLKRALGKFGREKKRMRÖRQAELEKVGGRVILFK 280

RESULT 3

US-08-683-458-6
; Sequence 6, Application US/08683458
; Patent No. 5798233
; GENERAL INFORMATION:
; APPLICANT: Gotschlich, Emil C.
; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
; NUMBER OF SEQUENCES: 12
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Klauder & Jackson
; STREET: 411 Hackensack Avenue
; CITY: Hackensack
; STATE: New Jersey
; COUNTRY: USA
; ZIP: 07601
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentin Release #1.0, Version #1.25
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/683,458
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 08/312,387
; FILING DATE: September 26, 1994
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Jackson Esq., David A.
; REFERENCE/DOCKET NUMBER: 26,742
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: 201 487-5800
; TELEFAX: 201 343-1684
; TELEX: 133521
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-683-458-6

Query Match 100.0%; Score 1440; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 7e-155;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MONHVISTLSAERRAHINDTFGSRGIPQFPDAMPSERLBOAMAEVPGLSAHPYLSG 60
Db 1 MONHVISTLSAERRAHINDTFGSRGIPQFPDAMPSERLBOAMAEVPGLSAHPYLSG 60
QY 61 VERACFMSHAVIWEQALDGLPIYIAVFEDVLLGGAEOFLAEDTWLEERFKDSAFIVR 120
Db 61 VERACFMSHAVIWEQALDGLPIYIAVFEDVLLGGAEOFLAEDTWLEERFKDSAFIVR 120
QY 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180
Db 121 LETMFAKVIVRDPKVLNENRSFPLESEHCCTAGYIISREAMRFFLDRAVLPERRIKA 180

Qy	181	VDLMMETFFPKKEMPMYQVSPALCTOELHAKRLSONSMGLSEKEDROGRHNRSLK	240
Db	181	VDLMMETFFPKKEMPMYQVSPALCTOELHAKRLSONSMGLSEKEDROGRHNRSLK	240
Qy	241	VMFDLKRALGKFGREKKKRMEDROQAELEKYGARVLLFK	280
Db	241	VMFDLKRALGKFGREKKKRMEDROQAELEKYGARVLLFK	280

RESULT 4

```

US-08-878-360-6
Sequence 6, Application US/08878360
Patent No. 5945322
GENERAL INFORMATION:
APPLICANT: Gotschlich, Emil C.
TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Klauber & Jackson
STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/878,360
FILING DATE: 18-JUN-1997
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/583,426
FILING DATE:
APPLICATION NUMBER: 08/312,387
FILING DATE: September 26, 1994
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095B
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 280 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-878-360-6

```

Query Match	100.0%	Score 1440	DB 2	Length 280
Best Local Similarity	100.0%	Pred. No. 7e-155		
Matches 280	0	Mismatches	0	Gaps 0

[illegible]

Oy	181	VDIIMFTYFEPKBEGRVQVSPALCTOELNIAKLSONSMGSDLEKDROGRNRHSLK	240
Dd	181	VDIIMFTYFEPKBEGRVQVSPALCTOELNIAKLSONSMGSDLEKDROGRNRHSLK	240
Oy	241	VFEDLKRALGKFGEKKRMRORQADELKYUGRRVLFLK	280
Dd	241	VFEDLKRALGKFGEKKRMRORQADELKYUGRRVLFLK	280

RESULT 5

```

US-08-478-140B-6
; Sequence 6, Application US/08478140B
; Patent No. 6127153
; GENERAL INFORMATION:
; APPLICANT: JOHNSON, KARL F.
; APPLICANT: ROTH, STEPHEN
; APPLICANT: BOCCALA, STEPHANIE L.
; TITLE OF INVENTION: METHOD OF TRANSFERRING AT LEAST TWO
; TITLE OF INVENTION: SACCHARIDE UNITS WITH A POLYGLYCOSYLTRANSFERASE, A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE AND GENE ENCODING A
; TITLE OF INVENTION: POLYGLYCOSYLTRANSFERASE
; NUMBER OF SEQUENCES: 8
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Pennie & Edmonds
; STREET: 1155 Avenue of the Americas
; CITY: New York
; STATE: New York
; COUNTRY: USA
; ZIP: 10036-2711
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Floppy disk
; OPERATING SYSTEM: IBM PC compatible
; SOFTWARE: Patentln Release #1.0, Version #1.30
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/478.140B
; FILING DATE: 07-JUN-1995
; CLASSIFICATION: 435
; ATTORNEY/AGENT INFORMATION:
; NAME: Laura A. Coruzzi
; REGISTRATION NUMBER: 30,742
; REFERENCE/DOCKET NUMBER: 7188-017
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (212) 790-9090
; TELEFAX: (212) 869-9741/8864
; TELEX: 66141 PENNIE
; INFORMATION FOR SEQ ID NO: 6:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 280 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-478-140B-6

```

Query Match	100.0%	Score 1440	DB 3	Length 280
Best Local Similarity	100.0%	Pred. No. 7e-155		
Matches 280	0	Mismatches	0	Gaps 0

QY	1	MONHVISLASAERRAHIAADTFPGSGHCPDPOFDALMPSESRLEQMAELVPGLSAHPIYLSG	60
Db	1	MONHVISLASAERRAHIAADTFPGSGHCPDPOFDALMPSESRLEQMAELVPGLSAHPIYLSG	60
QY	61	VEKAEFMSHAYLMEQALDDEGLPYIAVEFDDVLLLEBGAEOFLAEPDWLLEBRDKDSATIVR	120
Db	61	VEKAEFMSHAYLMEQALDDEGLPYIAVEFDDVLLLEBGAEOFLAEPDWLLEBRDKDSATIVR	120
QY	121	LETMAKAIYVPDKVLNENRSFPLLESEHOGTAGYIISREAMRFPLDRFAVLPERIKA	180
Db	121	LETMAKAIYVPDKVLNENRSFPLLESEHOGTAGYIISREAMRFPLDRFAVLPERIKA	180
QY	181	VDLMEFTYFDKEGMPYQVSPALCTOELHAYAKFLSONSMGLSDEKRDREGRRHRSLSK	240
Db	181	VDLMEFTYFDKEGMPYQVSPALCTOELHAYAKFLSONSMGLSDEKRDREGRRHRSLSK	240

Db 241 VWFDLKALGKFGREKKRMRORQAELEKYGRVILFK 280

RESULT 8

US-08-312-387B-8

; Sequence 8, Application US/08312387B

; Patent No. 5543553

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/312,387B

; FILING DATE: July 7, 1994

; CLASSIFICATION: 435

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-095

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201 487-5800

; TELEFAX: 201 343-1684

; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 279 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-312-387B-8

Query Match 72.0%; Score 1036.5; DB 1; Length 279;

Best Local Similarity 71.6%; Pred. No. 3.5e-109;

Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;

QY 1 MONHYISLASAERRAHADTFGSRGIPFOFPDALMPSERLERQAMAEIVPGISAHPIYLSG 60

DB 1 MONHYISLASAERRAHADTFGSRGIPFOFPDALMPSERLERQAMAEIVPGISAHPIYLSG 60

QY 61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAGEOFLAEDTWLEERFDKDSAFIVR 120

DB 61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAGEOFLAEDTWLEERFDKDSAFIVR 120

QY 121 LETMFAYIVRPDKYLVNENRSEFPLESEHCCTAGYIISREMRFFLDRAVLPERIKR 180

DB 121 LETMFAYIVRPDKYLVNENRSEFPLESEHCCTAGYIISREMRFFLDRAVLPERIKR 180

QY 181 VDLAMFTFFDEKGMPIYQVSPALCTOELHYAKFLSÖNSMLGSDLEKDRQGR----- 233

DB 181 VDLAMFTFFDEKGMPIYQVSPALCTOELHYAKFLSÖNSMLGSDLEKDRQGR----- 233

QY 234 -----RHRSRLKVMFMDLKRALGKFGREKKRMRORQAELEKYGRVILFK 280

DB 241 PANTFKHR-----LIRALTKIGRERKRRORR-----EDLIGKIIVPFO 279

RESULT 9

US-08-683-426-8

; Sequence 8, Application US/08683426

; Patent No. 5705367

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

; TITLE OF INVENTION: OLIGOSACCHARIDES, AND GENES ENCODING THEM

; NUMBER OF SEQUENCES: 12

; CORRESPONDENCE ADDRESS:

; ADDRESSEE: Klauber & Jackson

; STREET: 411 Hackensack Avenue

; CITY: Hackensack

; STATE: New Jersey

; COUNTRY: USA

; ZIP: 07601

; COMPUTER READABLE FORM:

; MEDIUM TYPE: Floppy disk

; COMPUTER: IBM PC compatible

; OPERATING SYSTEM: PC-DOS/MS-DOS

; SOFTWARE: Patentln Release #1.0, Version #1.25

; CURRENT APPLICATION DATA:

; APPLICATION NUMBER: US/08/683,426

; FILING DATE:

; CLASSIFICATION: 536

; PRIOR APPLICATION DATA:

; APPLICATION NUMBER: 08/312,387

; FILING DATE: September 26, 1994

; CLASSIFICATION: 536

; ATTORNEY/AGENT INFORMATION:

; NAME: Jackson Esq., David A.

; REGISTRATION NUMBER: 26,742

; REFERENCE/DOCKET NUMBER: 600-1-095B

; TELECOMMUNICATION INFORMATION:

; TELEPHONE: 201 487-5800

; TELEFAX: 201 343-1684

; TELEX: 133521

; INFORMATION FOR SEQ ID NO: 8:

; SEQUENCE CHARACTERISTICS:

; LENGTH: 279 amino acids

; TYPE: amino acid

; TOPOLOGY: linear

; MOLECULE TYPE: protein

US-08-683-426-8

Query Match 72.0%; Score 1036.5; DB 1; Length 279;

Best Local Similarity 71.6%; Pred. No. 3.5e-109;

Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;

QY 1 MONHYISLASAERRAHADTFGSRGIPFOFPDALMPSERLERQAMAEIVPGISAHPIYLSG 60

DB 1 MONHYISLASAERRAHADTFGSRGIPFOFPDALMPSERLERQAMAEIVPGISAHPIYLSG 60

QY 61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAGEOFLAEDTWLEERFDKDSAFIVR 120

DB 61 VEKACFMSHAYLMEQALDEGLPYIAVFEDDVLGGAGEOFLAEDTWLEERFDKDSAFIVR 120

QY 121 LETMFAYIVRPDKYLVNENRSEFPLESEHCCTAGYIISREMRFFLDRAVLPERIKR 180

DB 121 LETMFAYIVRPDKYLVNENRSEFPLESEHCCTAGYIISREMRFFLDRAVLPERIKR 180

QY 181 VDLAMFTFFDEKGMPIYQVSPALCTOELHYAKFLSÖNSMLGSDLEKDRQGR----- 233

DB 181 VDLAMFTFFDEKGMPIYQVSPALCTOELHYAKFLSÖNSMLGSDLEKDRQGR----- 233

QY 234 -----RHRSRLKVMFMDLKRALGKFGREKKRMRORQAELEKYGRVILFK 280

DB 241 PANTFKHR-----LIRALTKIGRERKRRORR-----EDLIGKIIVPFO 279

RESULT 10

US-08-683-458-8

; Sequence 8, Application US/08683458

; Patent No. 5798233

; GENERAL INFORMATION:

; APPLICANT: Gotschlich, Emil C.

; TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF

STREET: 411 Hackensack Avenue
CITY: Hackensack
STATE: New Jersey
COUNTRY: USA
ZIP: 07601
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/333,412
FILING DATE: 15-Jun-1999
CLASSIFICATION: <unknown>
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/312,387
FILING DATE: July 7, 1994
ATTORNEY/AGENT INFORMATION:
NAME: Jackson Esq., David A.
REGISTRATION NUMBER: 26,742
REFERENCE/DOCKET NUMBER: 600-1-095
TELECOMMUNICATION INFORMATION:
TELEPHONE: 201 487-5800
TELEFAX: 201 343-1684
TELEX: 133521
INFORMATION FOR SEQ ID NO: 8:
SEQUENCE CHARACTERISTICS:
LENGTH: 279 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
SEQUENCE DESCRIPTION: SEQ ID NO: 8:
US-09-333-412-8
Query Match 72.0%; Score 1036.5; DB 4; Length 279;
Best Local Similarity 71.6%; Pred. No. 3.3e-109;
Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;
QY 1 MONHYSLSAERRAHIAIDTFGSGIIPFPFDALMPSERLQAMAEVPGLSAHPYLSG 60
DB 1 MONHYSLSAERRAHIAIDTFGSGIIPFPFDALMPSERLQAMAEVPGLSAHPYLSG 60
QY 61 VEKACPMHVAIMEQALDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120
DB 61 VEKACPMHVAIMEQALDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120
QY 121 LETMAKVIYVRDQKLVNTNRSFPLSEHCCTAGYIISREMRFFLDRAVLPERRIKA 180
DB 121 LETMAKVIYVRDQKLVNTNRSFPLSEHCCTAGYIISREMRFFLDRAVLPERRIKA 180
QY 181 VDLMMETFEDEGMPVYOVSPALCTOEIHYAKPLSONSMGSDLEKDEKDEQGR----- 233
DB 181 VDLMMETFEDEGMPVYOVSPALCTOEIHYAKPLSONSMGSDLEKDEKDEQGR----- 233
QY 234 -----RHRRSLKVMFDLKRALGKFGREKKRMRERQRAELEKVGRRVILFK 280
DB 241 PANTFGRH-----LIRALFKIGREKREKRQR-----EQLGKIIIVFQ 279
RESULT 13
US-08-646-590B-40
Sequence 40, Application US/08646590B
Patent No. 5962283
GENERAL INFORMATION:
APPLICANT: Warren, Patrick V.
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:
ADDRESS: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA

COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FastSeq for Windows Version 2.0
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/646,590B
FILING DATE: 08-May-1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/599,171
FILING DATE: 09-FEB-1996
PRIOR APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01094
FILING DATE: 21-January-1997
CLASSIFICATION: 435
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/017001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO: 40:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-08-646-590B-40
Query Match 6.9%; Score 99; DB 2; Length 394;
Best Local Similarity 22.8%; Pred. No. 0.0084;
Matches 75; Conservative 38; Mismatches 126; Indels 90; Gaps 17;
QY 9 ASAAERRAHIAITFG-----SGIIP---FQFPA 34
DB 22 AKAKELRAKGVIVIGFAGEPDPDPDTFKENCIALRGKTKYVPSACIPELRAIAEK 81
QY 35 LMPSERLQAMAEVPGLSAHPYLSGVEKACPMHVAIMEQALDEG-----LPIAVFE 88
DB 82 LKKEKVEKPESEIYVS-----AGAKMVLF-----IFMALLDGDDEVLLSPYVWYTP 130
QY 89 DVLGEGAEQFLAEDTWLEERFDKDSAFIVRLETMAKVIYVRDK-VIN-----Y 138
DB 131 EQIRFEGV-----VEVPLKKEKGFOLSEDEVKEKVTERTKAIYVINSNNPTGAVY 182
QY 139 EKRSPFLSEHCCTAG-YIISREMRFFLDRAVLPERRIKAVDLMETFEFDPKE 193
DB 183 EEEELKTI-AEPCVERGIFITISDECYEYFVGDAKFVSPASDSDEVKNTTFVNAFSSSY 241
QY 194 GMPVYOVSPALCTOEIHYAKPL-SONSMGSDLEKDEKDEKRRHRSRLKVMFDLKRALGK- 251
DB 242 SMTGRRIGYVACPEE-YAKVYIASLNSQSVSVTTFQAQY-----ALELAKNPKSD 291
QY 252 FGREKKRMRERQRA-----ELEKVGRRVILFK 277
DB 292 FVNEKRNAFERRRRDRAVELSKIPGMDVY 320
RESULT 14
US-09-412-184-40
Sequence 40, Application US/09412184
Patent No. 6268188
GENERAL INFORMATION:
APPLICANT: Warren, Patrick V.
TITLE OF INVENTION: TRANSAMINASES AND AMINOTRANSFERASES
NUMBER OF SEQUENCES: 42
CORRESPONDENCE ADDRESS:

```

ADDRESS: Fish & Richardson, P.C.
STREET: 4225 Executive Square, Suite 1400
CITY: La Jolla
STATE: CA
COUNTRY: US
ZIP: 92037
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette
COMPUTER: IBM Compatible
OPERATING SYSTEM: Windows95
SOFTWARE: FASTSEQ for Windows
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/09/412,184
CLASSIFICATION:
FILING DATE:
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US/08/646,590
FILING DATE: 08-May-1996
APPLICATION NUMBER: 08/599,171
FILING DATE: 09-FEB-1996
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: PCT/US97/01094
FILING DATE: 21-January-1997
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Haile, Ph.D., Lisa A.
REGISTRATION NUMBER: 38,347
REFERENCE/DOCKET NUMBER: 09010/017001
TELECOMMUNICATION INFORMATION:
TELEPHONE: 619/678-5070
TELEFAX: 619/678-5099
INFORMATION FOR SEQ ID NO:
SEQUENCE CHARACTERISTICS:
LENGTH: 394 amino acids
TYPE: amino acid
TOPOLOGY: linear
MOLECULE TYPE: protein
FRAGMENT TYPE: internal
US-09-412-184-40

```

```

Query Match          6.9%; Score 99; DB 4; Length 394;
Best Local Similarity 22.8%; Pred. No. 0.0084;
Matches 75; Conservative 38; Mismatches 126; Indels 90; Gaps 17;

```

```

QY 9 ASAEERRAHIAIDTFG-----SRGIP---FQGFDA 34
DB 22 AKAKEIRAGGVYDIGAGBEPDPTDFIKACIRALRECKTKYASAGIPELREIAEK 81
QY 35 LMSERLEQAMAMELVPGLSAHPLYSGVERACFMSHAIVMEQALDEG-----LPYIAVEE 88
DB 82 LKKNKVEKSPSEIVS-----AGAKMYLFU---IFMAILDEGDEVLLPSPYVWTYP 130
QY 89 DVLVLEGAGDQFLAEDTWLEERDKDSAFIVLETFPAKYIYVPPDK-VLN-----Y 138
DB 131 EQIRFEGVAP-----EVLPLKKEKGFQLSLEDVKEKTEVETKTAIVNSPNPTGAVY 182
QY 139 ENNSFPLTSEHSGTAG-YIISREARFPL--DRFAVLPERIKAKVLDMMFTY--PFDE 193
DB 183 EEEELKRT-AECVGERGITIISDECEYTFYGDAKFVSPASISDEKNTITFYNAFSKY 241
QY 194 GMFVYVSPALCTQELHYAKFL-SQNSMLGSDLEKDEQGRHRRSLKMYFDLKRALGK- 251
DB 242 SMWGMIIGVACEE--YAKVIASLNSQSVSNVTTFAYG-----ALEALKNKNSMD 291
QY 252 FGREKKRMRORQA--ELEKYYGRRVI 277
DB 292 FVEMENRAFERRDYAVEELSKIPMDVY 320

```

```

RESULT 15
US-08-468-576B-17
; Sequence 17, Application US/08468576B
; Patent No. 5955345

```

```

GENERAL INFORMATION:
APPLICANT: Rabin, Daniel
TITLE OF INVENTION: PANCREATIC ISLET CELL ANTIGENS
TITLE OF INVENTION: OBTAINED BY MOLECULAR CLONING
NUMBER OF SEQUENCES: 19
CORRESPONDENCE ADDRESS:
ADDRESS: Sprung Kramer Schaefer & Briscoe
STREET: 660 White Plains Road
CITY: Tarrytown
STATE: New York
COUNTRY: USA
ZIP: 10591-5144
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4 Mb storage
COMPUTER: Apple Macintosh
OPERATING SYSTEM: System 7.5
SOFTWARE: Wordperfect
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/468,576B
FILING DATE: 06-JUN-1995
CLASSIFICATION: 435
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 08/239,276
FILING DATE: 05-MAY-1994
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/872,646
FILING DATE: 08-JUN-1992
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/715,181
FILING DATE: 14-JUN-1991
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/441,703
FILING DATE: 04-DEC-1989
PRIORITY APPLICATION DATA:
APPLICATION NUMBER: US 07/312,543
FILING DATE: 17-FEB-1989
ATTORNEY/AGENT INFORMATION:
NAME: Kuit G. Briscoe
REGISTRATION NUMBER: 33,141
REFERENCE/DOCKET NUMBER: MDI 251.7-KGB
TELECOMMUNICATION INFORMATION:
TELEPHONE: (914) 332-1700
TELEFAX: (914) 332-1844
INFORMATION FOR SEQ ID NO: 17:
SEQUENCE CHARACTERISTICS:
LENGTH: 712 amino acids
TYPE: amino acid
TOPOLOGY: linear
US-08-468-576B-17

```

```

Query Match          5.8%; Score 83; DB 2; Length 712;
Best Local Similarity 20.2%; Pred. No. 1.3;
Matches 66; Conservative 59; Mismatches 106; Indels 96; Gaps 17;

```

```

QY 5 VISLSAERRAHIAIDTFGSRGIPFQFEDALMPSERLEQAMAMELVPGLSAHPLYSG----- 60
DB 10 VADVLSQDLRKRIKIVQTFMDRGI-----IDTLKTQLR-NQLIHLELM-----HYVLSGELQ 59
QY 61 ----VEKACFM--SHAVIMEQALDEGLPY-IAVEDVVLGEGADQFLAEDTWLEERPD 112
DB 60 RSIYSVEGSLILGASNSLVADHLQRCGYEYSLSVF-----PFESGIA 101
QY 113 KDSAFIVR-----LETMFAYIVPPDKVLYNENRSPFL----- 145
DB 102 KEKVFMTQDLQILKINPTSLYKSLVSGSDK-----ENOKGFLMFLKELAEYHOAKESC 157
QY 146 -LESEHCGTAGYIISREARFPLDRFAVLPERIK--AVDLMMFTYFPDEKGPYVYQVSP 202
DB 158 NMEIOTSSFTFNDSLAERKQLIDDOFADAYPQRIKTFESLEILNYYKKRIE-----EQLR 213
QY 203 ALCTQELHYAKFLSONSMLGSDLEKDEQGRHRRSLKMYF--DLKRALGKFG-----RE 255
DB 214 EMC-QKIKPFK-----DTETAKIKMEAKKKYKEKELTMFONDEKACQAKSEALVIRE 264

```


Oy 256 KK-KRMRORQAELEKVGRRVILEK 280
| : : : | : : | : |
Db 265 KSTLERIHKHQEIETKTEIYAQRQLLK 291

Search completed: December 2, 2002, 12:03:18
Job time : 13.4223 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 12:00:15 ; Search time 5.61203 Seconds

(without alignments)
794.504 Million cell updates/sec

Title: US-10-007-267-6

Perfect score: 1440

Sequence: 1 MGNHVISLASAERAHIAID.....ERQRAELEKVGRRVILFK 280

Scoring table: BLOSUM62

Gapop 10.0 , Gapext 0.5

Searched: 102317 seqs, 15924203 residues

Total number of hits satisfying chosen parameters: 102317

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%

Listing first 45 summaries

Database : Published_Applications_AA:*

1: /cgn2_6/ptodata/1/pubpaa/US08_NEM_PUB.pep:*

2: /cgn2_6/ptodata/1/pubpaa/PC1_NEM_PUB.pep:*

3: /cgn2_6/ptodata/1/pubpaa/US06_NEM_PUB.pep:*

4: /cgn2_6/ptodata/1/pubpaa/US06_PUBCOMB.pep:*

5: /cgn2_6/ptodata/1/pubpaa/US07_NEM_PUB.pep:*

6: /cgn2_6/ptodata/1/pubpaa/US07_PUBCOMB.pep:*

7: /cgn2_6/ptodata/1/pubpaa/PC1US_PUBCOMB.pep:*

8: /cgn2_6/ptodata/1/pubpaa/US08_PUBCOMB.pep:*

9: /cgn2_6/ptodata/1/pubpaa/US09_NEM_PUB.pep:*

10: /cgn2_6/ptodata/1/pubpaa/US09_PUBCOMB.pep:*

11: /cgn2_6/ptodata/1/pubpaa/US10_NEM_PUB.pep:*

12: /cgn2_6/ptodata/1/pubpaa/US10_PUBCOMB.pep:*

13: /cgn2_6/ptodata/1/pubpaa/US60_NEM_PUB.pep:*

14: /cgn2_6/ptodata/1/pubpaa/US60_PUBCOMB.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1440	100.0	280	US-10-007-267-6	Sequence 6, Appl1
2	1036.5	72.0	279	US-10-007-267-8	Sequence 8, Appl1
3	1007	69.9	275	US-09-211-691-2	Sequence 2, Appl1
4	999	6.9	294	US-09-905-173-40	Sequence 40, Appl1
5	87.5	6.1	657	US-08-815-242-5280	Sequence 5280, Ap
6	87.5	6.1	662	US-08-815-242-12140	Sequence 12140, A
7	85	5.9	762	US-08-815-242-11519	Sequence 11519, A
8	83.5	5.8	856	US-09-815-242-11310	Sequence 11310, A
9	79	5.5	2055	US-10-017-216-4	Sequence 4, Appl1
10	77.5	5.4	222	US-09-864-761-48086	Sequence 48086, A
11	77.5	5.4	363	US-09-849-031A-1	Sequence 1, Appl1
12	77.5	5.4	363	US-09-849-562A-1	Sequence 1, Appl1
13	77.5	5.4	623	US-09-815-242-13499	Sequence 13499, A
14	77.5	5.4	1224	US-09-801-368-222	Sequence 222, App
15	77.5	5.4	4999	US-09-976-059-14	Sequence 14, Appl1
16	77	5.3	592	US-09-861-451A-72	Sequence 72, Appl1
17	76.5	5.3	856	US-09-815-242-11489	Sequence 11489, A
18	76.5	5.3	916	US-09-745-763-174	Sequence 174, App
19	76.5	5.3	1005	US-09-925-301-1335	Sequence 1335, Ap

20	75.5	5.2	989	US-09-815-242-4897	Sequence 4897, Ap
21	74	5.1	302	US-09-815-242-5371	Sequence 5371, Ap
22	74	5.1	325	US-09-815-242-12608	Sequence 12608, A
23	73.5	5.1	883	US-09-815-242-13382	Sequence 13382, A
24	73.5	5.1	1427	US-09-991-496-97	Sequence 97, Appl1
25	73.5	5.1	1427	US-09-874-923-97	Sequence 97, Appl1
26	73.5	5.1	1641	US-09-991-496-96	Sequence 96, Appl1
27	73.5	5.1	1641	US-09-874-923-96	Sequence 96, Appl1
28	73	5.1	474	US-09-815-242-5389	Sequence 5389, A
29	73	5.1	474	US-09-815-242-12656	Sequence 12656, A
30	73	5.1	3014	US-09-737-149-2	Sequence 2, Appl1
31	72.5	5.0	556	US-09-887-586A-32	Sequence 32, Appl1
32	72.5	5.0	556	US-09-903-012-32	Sequence 32, Appl1
33	72.5	5.0	1237	US-09-841-132-592	Sequence 592, App
34	72.5	5.0	1579	US-09-801-368-368	Sequence 368, App
35	72	5.0	464	US-09-902-941-1906	Sequence 1906, Ap
36	72	5.0	468	US-09-925-300-1620	Sequence 1620, Ap
37	72	5.0	793	US-09-815-242-13689	Sequence 13689, A
38	72	5.0	1356	US-09-801-368-306	Sequence 306, App
39	71.5	5.0	99	US-09-864-761-33635	Sequence 33635, A
40	71.5	5.0	552	US-09-817-764-4	Sequence 4, Appl1
41	71.5	5.0	883	US-09-815-242-13684	Sequence 13684, A
42	71	4.9	420	US-09-844-006A-2	Sequence 2, Appl1
43	71	4.9	710	US-09-815-242-10895	Sequence 10895, A
44	71	4.9	734	US-09-764-367A-4	Sequence 4, Appl1
45	71	4.9	893	US-09-916-790-5	Sequence 5, Appl1

ALIGNMENTS

RESULT 1

US-10-007-267-6

; Sequence 6, Application US/10007267

; Patent No. US20020127682A1

GENERAL INFORMATION:

APPLICANT: Gotschlich, Emil C.

TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF OLIGOSACCHARIDES, AND GENES ENCODING THEM

NUMBER OF SEQUENCES: 12

CORRESPONDENCE ADDRESSES:

ADDRESSEE: Klauber & Jackson

STREET: 411 Hackensack Avenue

CITY: Hackensack

STATE: New Jersey

COUNTRY: USA

ZIP: 07601

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk

COMPUTER: IBM PC compatible

OPERATING SYSTEM: PC-DOS/MS-DOS

SOFTWARE: Patentin Release #1.0, Version #1.25

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/10/007,267

FILING DATE: 03-Dec-2001

CLASSIFICATION: <Unknown>

PRIOR APPLICATION DATA:

APPLICATION NUMBER: US/09/333,412

FILING DATE: 15-Jun-1999

APPLICATION NUMBER: 08/312,387

FILING DATE: July 7, 1994

ATTORNEY/AGENT INFORMATION:

NAME: Jackson Esq., David A.

REGISTRATION NUMBER: 26,742

TELECOMMUNICATION INFORMATION:

TELEPHONE: 201 487-5800

TELEFAX: 201 343-1684

TELEX: 133521

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 280 amino acids

TYPE: amino acid

```
;
;      TOPOLOGY: linear
;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 6:
us-10-007-267-6

Query Match          100.0%; Score 1440; DB 12; Length 280;
Best Local Similarity 100.0%; Pred. No. 2,3e-138;
Matches 280; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 MOUNHVISLASAARRRAHIAITFGSRGIPFOFPALMPSERLBOAMALVPGLSAHPYLSG 60
DB 1 MOUNHVISLASAARRRAHIAITFGSRGIPFOFPALMPSERLBOAMALVPGLSAHPYLSG 60
QY 61 VERACFMSHAIVLMEQALDEGLPIYAVEEDVLLGEGAEQFLADDTMLEERFDKDSAFIVR 120
DB 61 VERACFMSHAIVLMEQALDEGLPIYAVEEDVLLGEGAEQFLADDTMLEERFDKDSAFIVR 120
QY 121 LETMFAKVIYRPDKVLNENRSFPLLESEHCGTAGYIISREARFFLDRAVLPPERIKR 180
DB 121 LETMFAKVIYRPDKVLNENRSFPLLESEHCGTAGYIISREARFFLDRAVLPPERIKR 180
QY 181 VDLMMFTYFDDKGMPPYQVSPALCTQOELHYAKFLSONSMGLSDLEKDRQGRHRSLSK 240
DB 181 VDLMMFTYFDDKGMPPYQVSPALCTQOELHYAKFLSONSMGLSDLEKDRQGRHRSLSK 240
QY 241 VMFDLKRALGKFGREKKRMRERORQAELEKYGRVILFK 280
DB 241 VMFDLKRALGKFGREKKRMRERORQAELEKYGRVILFK 280

RESULT 2
US-10-007-267-8
; Sequence 8, Application US/10007267
; Patent No. US20020127682A1
; GENERAL INFORMATION:
;   APPLICANT: Gotschlich, Emil C.
;   TITLE OF INVENTION: GLYCOSYLTRANSFERASES FOR BIOSYNTHESIS OF
;                     OLIGOSACCHARIDES, AND GENES ENCODING THEM
;   NUMBER OF SEQUENCES: 12.
;   CORRESPONDENCE ADDRESS:
;     ADDRESSER: Klauber & Jackson
;     STREET: 411 Hackensack Avenue
;     CITY: Hackensack
;     STATE: New Jersey
;     COUNTRY: USA
;     ZIP: 07601
;   COMPUTER READABLE FORM:
;     COMPUTER: IBM PC compatible
;     OPERATING SYSTEM: PC-DOS/MS-DOS
;     SOFTWARE: PatentIn Release #1.0, Version #1.25
;   CURRENT APPLICATION DATA:
;     APPLICATION NUMBER: US/10/007,267
;     FILING DATE: 03-Dec-2001
;     CLASSIFICATION: <Unknown>
;   PRIOR APPLICATION DATA:
;     APPLICATION NUMBER: US/09/333,412
;     FILING DATE: 15-Jun-1999
;     APPLICATION NUMBER: 08/312,387
;     FILING DATE: July 7, 1994
;   ATTORNEY/AGENT INFORMATION:
;     NAME: Jackson Esq., David A.
;     REGISTRATION NUMBER: 26,742
;     REFERENCE/DOCKET NUMBER: 600-1-095
;   TELECOMMUNICATION INFORMATION:
;     TELEPHONE: 201 487-5800
;     TELEFAX: 201 343-1684
;     TELE: 133521
;   INFORMATION FOR SEQ ID NO: 8:
;     SEQUENCE CHARACTERISTICS:
;       LENGTH: 279 amino acids
;       TYPE: amino acid
;       TOPOLOGY: linear
```

```
;
;      MOLECULE TYPE: protein
;      SEQUENCE DESCRIPTION: SEQ ID NO: 8:
us-10-007-267-8

Query Match          72.0%; Score 1036.5; DB 12; Length 279;
Best Local Similarity 71.6%; Pred. No. 1.6e-97;
Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;

QY 1 MOUNHVISLASAARRRAHIAITFGSRGIPFOFPALMPSERLBOAMALVPGLSAHPYLSG 60
DB 1 MOUNHVISLASAARRRAHIAITFGSRGIPFOFPALMPSERLBOAMALVPGLSAHPYLSG 60
QY 61 VERACFMSHAIVLMEQALDEGLPIYAVEEDVLLGEGAEQFLADDTMLEERFDKDSAFIVR 120
DB 61 VERACFMSHAIVLMEQALDEGLPIYAVEEDVLLGEGAEQFLADDTMLEERFDKDSAFIVR 120
QY 121 LETMFAKVIYRPDKVLNENRSFPLLESEHCGTAGYIISREARFFLDRAVLPPERIKR 180
DB 121 LETMFAKVIYRPDKVLNENRSFPLLESEHCGTAGYIISREARFFLDRAVLPPERIKR 180
QY 181 VDLMMFTYFDDKGMPPYQVSPALCTQOELHYAKFLSONSMGLSDLEKDRQGR----- 233
DB 181 VDLMMFTYFDDKGMPPYQVSPALCTQOELHYAKFLSONSMGLSDLEKDRQGR----- 233
QY 234 -----RHRSLSKVMFDLKRALGKFGREKKRMRERORQAELEKYGRVILFK 280
DB 241 PANTFKHR-----LIRALTKIGREKRRRR-----EQLIGKIYVFG 279

RESULT 3
US-09-211-691-2
; Sequence 2, Application US/09211691
; Patent No. US20020034805A1
; GENERAL INFORMATION:
;   APPLICANT: Gilbert, Michel
;   APPLICANT: Young, N. Martin
;   APPLICANT: Wakarchuk, Warren W.
;   APPLICANT: National Research Council of Canada
;   TITLE OF INVENTION: Fusion Proteins for Use in Enzymatic Synthesis of
;   TITLE OF INVENTION: Oligosaccharides
;   FILE REFERENCE: 019957-012910US
;   CURRENT APPLICATION NUMBER: US/09/211,691
;   CURRENT FILING DATE: 1998-12-14
;   PRIOR APPLICATION NUMBER: US 60/069,443
;   PRIOR FILING DATE: 1997-12-15
;   NUMBER OF SEQ ID NOS: 18
;   SOFTWARE: PatentIn Ver. 2.1
;   SEQ ID NO 2
;   LENGTH: 275
;   TYPE: PRT
;   ORGANISM: Neisseria meningitidis
US-09-211-691-2

Query Match          69.9%; Score 1007; DB 10; Length 275;
Best Local Similarity 74.7%; Pred. No. 1.5e-94;
Matches 201; Conservative 20; Mismatches 44; Indels 4; Gaps 1;

QY 1 MOUNHVISLASAARRRAHIAITFGSRGIPFOFPALMPSERLBOAMALVPGLSAHPYLSG 60
DB 1 MOUNHVISLASAARRRAHIAITFGSRGIPFOFPALMPSERLBOAMALVPGLSAHPYLSG 60
QY 61 VERACFMSHAIVLMEQALDEGLPIYAVEEDVLLGEGAEQFLADDTMLEERFDKDSAFIVR 120
DB 61 VERACFMSHAIVLMEQALDEGLPIYAVEEDVLLGEGAEQFLADDTMLEERFDKDSAFIVR 120
QY 121 LETMFAKVIYRPDKVLNENRSFPLLESEHCGTAGYIISREARFFLDRAVLPPERIKR 180
DB 121 LETMFAKVIYRPDKVLNENRSFPLLESEHCGTAGYIISREARFFLDRAVLPPERIKR 180
QY 181 VDLMMFTYFDDKGMPPYQVSPALCTQOELHYAKFLSONSMGLSDLEKDRQGR----- 236
DB 181 VDLMMFTYFDDKGMPPYQVSPALCTQOELHYAKFLSONSMGLSDLEKDRQGR----- 240
```

OY 237 RSLKVPEDIKRALGKFGREKKRMRORQ 265
 Db 241 PANTFKHLRALTKISREKRORRQ 269

RESULT 4

US-09-905-173-40
 ; Sequence 40, Application US/09905173
 ; Patent No. US2002013295A1

GENERAL INFORMATION:

APPLICANT: DIVERSA CORPORATION
 APPLICANT: SHORT, Jay M.
 APPLICANT: WARREN, Patrick V.
 APPLICANT: SWANSON, Ronald V.
 APPLICANT: MATHUR, Eric J.
 TITLE OF INVENTION: ENZYMES HAVING TRANSAMINASE AND AMINOTRANSFERASE ACTIVITY AND MET
 TITLE OF INVENTION: USE THEREOF
 FILE REFERENCE: DIVER1240-7
 CURRENT APPLICATION NUMBER: US/09/905,173
 CURRENT FILING DATE: 2001-07-12
 PRIOR APPLICATION NUMBER: US 09/412,184
 PRIOR FILING DATE: 1999-10-04
 PRIOR APPLICATION NUMBER: US 09/389,537
 PRIOR FILING DATE: 1999-09-02
 PRIOR APPLICATION NUMBER: US 08/646,590
 PRIOR FILING DATE: 1996-05-08
 PRIOR APPLICATION NUMBER: US 08/599,171
 PRIOR FILING DATE: 1996-02-09
 PRIOR APPLICATION NUMBER: US 09/481,733
 PRIOR FILING DATE: 2000-01-11
 PRIOR APPLICATION NUMBER: US 09/069,226
 PRIOR FILING DATE: 1998-04-27
 NUMBER OF SEQ ID NOS: 40
 SOFTWARE: PatentIn version 3.0

SEQ ID NO 40
 LENGTH: 394
 TYPE: PRT
 ORGANISM: Aquifex
 US-09-905-173-40

Query Match 6.9%; Score 99; DB 10; Length 394;
 Best Local Similarity 22.8%; Pred. No. 0.021;
 Matches 75; Conservative 38; Mismatches 126; Indels 90; Gaps 17;

OY 9 ASAAERRAHIIADTFG-----SRGIP---FGFPA 34
 Db 22 AAKAKELRAGGVYIGFGAGEPDPTDFIKACIRALREGKTKYAPASGIPELREAIK 81
 OY 35 LMPSERLQAMAEIVPGLSAHPYLSGVEKACPMASHAVLMEQALDEG-----LPYIAVEE 88
 Db 82 LTKENKVEYKPSSEIVVS-----AGAKMVLFL---IFMAILDEGDEVLLSPYWTYP 130
 OY 89 DQVLGEGMGOFLAEDTWMLEERFDKDSAFIVRLTMAFVIRPK-VLN-----Y 138
 Db 131 EDIRFEGVP-----VEVPLKKEKGFQLSLEDVKEVTEKTAIVINSPNPTGAVY 182
 OY 139 ENRSFPLESEHQGTAG-YIISREARFPL--DRFAVLPPERIKAVDLAMFTY--FFDKE 193
 Db 183 EEBELKKT-AEFCVGERGITIISDECEYFYVGDAKFVSASDSDEKNTITFYNAFSKY 241
 OY 194 GMPYQVSPALCTQELHYAKFL-SQNSMLGSDLEKDEGRHRRSLKWMFDIKRALGK- 251
 Db 242 SMWGRIIGVACPEE--YAKVIASLNSQSVSVNTTFAQYG-----ALEALKNPKSKD 291
 OY 252 FGREKKRMRORQ---ELEKYGGRRVI 277
 Db 292 FVNEMRPAERRRDVAEELSKIPGMDVY 320

RESULT 5
 US-09-815-242-5280
 ; Sequence 5280, Application US/09815242
 ; Patent No. US20020061569A1

GENERAL INFORMATION:

APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes In
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A
 CURRENT APPLICATION NUMBER: US/09/815,242
 CURRENT FILING DATE: 2001-03-21
 PRIOR APPLICATION NUMBER: 60/191,078
 PRIOR FILING DATE: 2000-03-21
 PRIOR APPLICATION NUMBER: 60/206,848
 PRIOR FILING DATE: 2000-05-23
 PRIOR APPLICATION NUMBER: 60/207,727
 PRIOR FILING DATE: 2000-05-26
 PRIOR APPLICATION NUMBER: 60/242,578
 PRIOR FILING DATE: 2000-10-23
 PRIOR APPLICATION NUMBER: 60/253,625
 PRIOR FILING DATE: 2000-11-27
 PRIOR APPLICATION NUMBER: 60/257,931
 PRIOR FILING DATE: 2000-12-22
 PRIOR APPLICATION NUMBER: 60/269,308
 PRIOR FILING DATE: 2001-02-16
 NUMBER OF SEQ ID NOS: 14110
 SOFTWARE: FastSeq for Windows Version 4.0

SEQ ID NO 5280
 LENGTH: 657
 TYPE: PRT
 ORGANISM: Staphylococcus aureus
 US-09-815-242-5280

Query Match 6.1%; Score 87.5; DB 10; Length 657;
 Best Local Similarity 23.1%; Pred. No. 0.61;
 Matches 54; Conservative 22; Mismatches 79; Indels 79; Gaps 7;

OY 29 FQFPDLMPSERLEQMA-----ELVGLSAHPYLSGVEA 64
 Db 429 FVFSIDLKPKALRLSSIMGNATFTFTHDSIANGEDGPTHEPIQLAGLAIIPMNVIRPA 488
 OY 65 CFMSHAVLMEQALD-----EGLPYIAVFEDV----- 91
 Db 489 DGNETVAVNEVALLESSTPTSLVLTRONLPLVDYDPEDVYBEGYRKATYVYSGEETPEFL 548
 OY 92 LLEGAGEOFLAEDTWMLEERFDKDSAFIVRLTMAF-----KVIKPKVLYNE--NR 141
 Db 549 LLAGSGEVLAVEAAKDLKQKSVAVVSPMNNAFEGQSEYKESVIPPSTKRVAIEM 608
 OY 142 SFPLESEHCAGTAGYIISREARFPLDRFAVLPPERIKAVDLAMFTYFFPKDKGM 195
 Db 609 ASPLGMHKYVGTAGKIVA-----IDGFGASAPG-----DLAVEKYGPTKENI 650

RESULT 6
 US-09-815-242-12140
 ; Sequence 12140, Application US/09815242
 ; Patent No. US20020061569A1
 GENERAL INFORMATION:
 APPLICANT: Haselbeck, Robert
 APPLICANT: Ohlsen, Karl L.
 APPLICANT: Zyskind, Judith W.
 APPLICANT: Wall, Daniel
 APPLICANT: Trawick, John D.
 APPLICANT: Carr, Grant J.
 APPLICANT: Yamamoto, Robert T.
 APPLICANT: Xu, H. Howard
 TITLE OF INVENTION: Identification of Essential Genes In
 TITLE OF INVENTION: Prokaryotes
 FILE REFERENCE: ELITRA.011A

```
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ CURRENT FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 12140
/ LENGTH: 662
/ TYPE: PRT
/ ORGANISM: Staphylococcus aureus
US-09-815-242-12140
```

```
Query Match          6.1%; Score 87.5; DB 10; Length 662;
Best Local Similarity 23.1%; Pred. No. 0.62;
Matches 54; Conservative 22; Mismatches 79; Indels 79; Gaps 7;
```

```
QY 29 FGFPLALMSELEQAMA-----ELVYGLAHPLSLVEKA 64
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 434 FVPSDLKPLKRISSTIGLNATFTHTDSIANGEDGPTHEPIQLAGLRATPMNVIRPA 493
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 65 CFMSHAVLWEQALD-----EGLPYIAVEEDVY----- 91
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 494 DGNFTVAVAEVLESSTPTSLVLTFRQNLPLVDVPEDEVVEGVKRAYVYGESETPERL 553
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 92 LAGEGAQFLAEDPTLEERPKDSAFIVRLTFYA-----KVIYRPDKVLYNF--NR 141
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 554 LLASGEVSLAAVEAKDKCKSVRVGMPWMAFEQOSEEKESVYKRVAIM 613
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 142 SPPLESEHCAGAGVYISREARFPLDRFAVLPPERIKAVDLMFYFFDKEM 195
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 614 ASPLGKHKYVGTAGKYA-----IDFGASAPG-----DLVERKYGTPTKEMI 655
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

RESULT 7

```
US-09-815-242-11519
/ Sequence 11519, Application US/09815242
/ Patent No. US2002061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ PRIOR FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
```

```
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 11519
/ LENGTH: 762
/ TYPE: PRT
/ ORGANISM: Helicobacter pylori
US-09-815-242-11519
```

```
Query Match          5.9%; Score 85; DB 10; Length 762;
Best Local Similarity 25.0%; Pred. No. 1.3;
Matches 53; Conservative 32; Mismatches 65; Indels 62; Gaps 12;
```

```
QY 89 DVLLEGGAQFLA--BDTLEERPKDSAFIV-----RLTFPAKYIVPDKVINTENS 142
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 412 DELEGTDADEASSLXKTLLEKRLKQNNQIVYTHHKLRLSVMA-----ENKE 459
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 143 FPLL-----ESBH-----CGAGYIISRE--AMREFLDRFAVLPPERIKAVDLMFTYF 189
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 460 VELLAALYDEEKRPYTYTLKGYIGKSYAFETALRTGVFPFL--EKAKP-----YG 510
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 190 FKEGAPVYQVSPALCTOELHYAKFLSQNSMLGSDLEKDRQGRH--RSLKVMEDLKRA 248
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 511 EDERKRLN-----LIENS-----STLERELKQNEHLNALKEQEDLKNA 550
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
QY 249 LKFGREKKRMRQQALEKVGRRVILFK 280
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
DB 551 -WLEMERKQKEIFHHRKLELERSYQOALNLK 581
    |||||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
```

RESULT 8

```
US-09-815-242-11310
/ Sequence 11310, Application US/09815242
/ Patent No. US2002061569A1
/ GENERAL INFORMATION:
/ APPLICANT: Haselbeck, Robert
/ APPLICANT: Ohlsen, Karl L.
/ APPLICANT: Zyskind, Judith W.
/ APPLICANT: Wall, Daniel
/ APPLICANT: Trawick, John D.
/ APPLICANT: Carr, Grant J.
/ APPLICANT: Yamamoto, Robert T.
/ APPLICANT: Xu, H. Howard
/ TITLE OF INVENTION: Identification of Essential Genes in
/ FILE REFERENCE: ELITRA.011A
/ CURRENT APPLICATION NUMBER: US/09/815,242
/ PRIOR FILING DATE: 2001-03-21
/ PRIOR APPLICATION NUMBER: 60/191,078
/ PRIOR FILING DATE: 2000-03-21
/ PRIOR APPLICATION NUMBER: 60/206,848
/ PRIOR FILING DATE: 2000-05-23
/ PRIOR APPLICATION NUMBER: 60/207,727
/ PRIOR FILING DATE: 2000-05-26
/ PRIOR APPLICATION NUMBER: 60/242,578
/ PRIOR FILING DATE: 2000-10-23
/ PRIOR APPLICATION NUMBER: 60/253,625
/ PRIOR FILING DATE: 2000-11-27
/ PRIOR APPLICATION NUMBER: 60/257,931
/ PRIOR FILING DATE: 2000-12-22
/ PRIOR APPLICATION NUMBER: 60/269,308
/ PRIOR FILING DATE: 2001-02-16
/ NUMBER OF SEQ ID NOS: 14110
/ SOFTWARE: FASTSEQ for Windows Version 4.0
/ SEQ ID NO 11310
/ LENGTH: 856
/ TYPE: PRT
/ ORGANISM: Helicobacter pylori
US-09-815-242-11310
```

Query	Match	Similarity	5.8%	Score	83.5%	DB	10%	Length	856;
Best Local	Similarity	21.7%	Pred.	No. 2.2.					
Matches	54;	Conservative	40;	Mismatches	98;	Indels	57;	Gaps	12;
QY	39	ERLQOAMAEVPGSAHFYLSGVKACFMSHAVLMEQALDEGLPIYAFVEDVLGE---	95						
Db	255	ERLKVIVIEVKKSNVILFIDEI-----HTIYGAGSBEQMDAANLTKALARGELHT	307						
QY	96	GABQFLAEDTWLEE---REDKDSAFVIRETFMAKVIYPPDKVLNENRSEPL-----	145						
Db	308	IGA-----TTLEKRYKRYFEKDMA---LORRQPIILNPESTI---NEALQILRLGIKE	352						
QY	146	LESEH---OCTAGYIISRE-AMRPFIDRFVAVLPERIKADVLMMFTYFPKDEGPPYQV	200						
Db	353	TELEHNTTINDSALLISAKLSRITDRF--LPKALIDLDEGAALQIKWMESEPAKLS	410						
QY	201	SPALCTOELHAKFELSONSMIGSDLEKDEQRRHRSILKYMFDLKRALGKFGREKKRM	260						
Db	411	SVKSSIGRLMEK-----QALEMEKESNAKRMQELIKELSDLK-----EKKIQL	455						
QY	261	EROROAELE 269							
Db	456	EAQFENEKE 464							
RESULT 9									
US-10-017-216-4									
Sequence 4, Application US/10017216									
Patent No. US20020160483A1									
GENERAL INFORMATION:									
APPLICANT: KAPILLER-LIBERMANN, Rosana									
TITLE OF INVENTION: I3245, A No. US20020160483A1el Human Myotonic Dystrophy Type Prod									
TITLE OF INVENTION: Kinase and Uses Therefor									
FILE REFERENCE: 10147-5701									
CURRENT APPLICATION NUMBER: US/10/017, 216									
PRIOR APPLICATION NUMBER: 2001-10-23									
PRIOR FILING DATE: 2000-10-23									
NUMBER OF SEQ ID NOS: 7									
SOFTWARE: PatentIn Ver. 2.1									
SEQ ID NO 4									
LENGTH: 2055									
TYPE: PRT									
ORGANISM: Mus musculus									
US-10-017-216-4									
Query Match			5.5%	Score	79;	DB	9;	Length	2055;
Best Local Similarity			17.1%	Pred.	No. 21;				
Matches	49;	Conservative	55;	Mismatches	99;	Indels	84;	Gaps	12;
QY	29	FOFF-----DALMPSEERLOAMAEI-----VGLSAHFYLSGVKACFMSHAVLMEQ	75						
Db	317	FQRFLEKFPDPKVSSELLDLQSLCVCYKERRKPFGLCCHFPFARTD-----WNN	366						
QY	76	ALDEGLPIYAFVEDVLGEBAEQFLAEDTWLEERFDKDSAFVIRETFMAKVIYPPDKV	135						
Db	367	IRNSPPPEVPTLKS-----DDT---SNFDEPE-----KSMWAFILCVPAEP	405						
QY	136	LNENRSPPLLESEHCGTAGYIISREAMRPFIDRFVAVLPERIKAVDLMF-----	186						
Db	406	LAFSEELLPFGFYSKALGYIGRSESYSVSSLD-----SPAKVSSMEKKLLIKSELODS	460						
QY	187	---TYFPFKE-----GMPYQVSPALCTOELHAKFELSONSMIGSD-----LEKDR	229						
Db	461	ODCKIKMQDEMTRLHRRYSEVAVLSQKEVELKASETOBSLIEDDLATYITECSLKSL	520						
QY	230	FOGR-----RRRSILKVMFDLKRALGKFGREKKRMEROROAELKRV	271						
Db	521	EQARMEVSOEDDKALQLLHDIREQ-----SRKLQETKEQDEYQAQVDEM	563						

```

Sequence 4866, Application US/09864761
Patent No. US20020048763A1
GENERAL INFORMATION:
APPLICANT: Penn, Sharon G.
APPLICANT: Rank, David R.
APPLICANT: Hanzel, David K.
APPLICANT: Chen, Wensheng
TITLE OF INVENTION: HUMAN GENOME-DERIVED SINGLE EXON NUCLEIC ACID PROBES USEFUL FOR
FILE REFERENCE: Aecomica-X-1
CURRENT APPLICATION NUMBER: US/09/864,761
CURRENT FILING DATE: 2001-05-23
PRIORITY APPLICATION NUMBER: US 60/180,312
PRIORITY FILING DATE: 2000-02-04
PRIORITY APPLICATION NUMBER: US 60/207,456
PRIORITY FILING DATE: 2000-05-26
PRIORITY APPLICATION NUMBER: US 09/632,366
PRIORITY FILING DATE: 2000-08-03
PRIORITY APPLICATION NUMBER: GB 24263.6
PRIORITY FILING DATE: 2000-10-04
PRIORITY APPLICATION NUMBER: US 60/236,359
PRIORITY FILING DATE: 2000-09-27
PRIORITY APPLICATION NUMBER: PCT/US01/00666
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00667
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00664
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00669
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00665
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00668
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00663
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00662
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00661
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: PCT/US01/00670
PRIORITY FILING DATE: 2001-01-30
PRIORITY APPLICATION NUMBER: US 60/234,687
PRIORITY FILING DATE: 2000-09-21
PRIORITY APPLICATION NUMBER: US 09/608,408
PRIORITY FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/774,203
PRIORITY FILING DATE: 2001-01-29
NUMBER OF SEQ ID NOS: 49117
SOFTWARE: Anomcam Sequence Listing Engine vers. 1.1
SEQ ID NO 48086
LENGTH: 222
TYPE: PRT
ORGANISM: Homo sapiens
FEATURE:
OTHER INFORMATION: MAP TO AC005532.1
OTHER INFORMATION: EXPRESSED IN LUNG, SIGNAL = 1.5
OTHER INFORMATION: EXPRESSED IN PLACENTA, SIGNAL = 1.5
OTHER INFORMATION: SWISSPROT HIT: Q50950, EVALUATE 4.90e-01
OTHER INFORMATION: EST_HUMAN HIT: BE336820.1, EVALUATE 1.00e-112
US-09-864-761-48086

Query Match          5.4%; Score 77.5; DB 10; Length 222;
Best Local Similarity 25.7%; Pred. No. 1.5;
Matches 19; Conservative 19; Mismatches 33; Indels 3; Gaps 2;

QY      98 BOFLAEDTWLEBERPKDSAFIVRLTMEFAKVIYRPDKVLYNENRSPFLESSEH-CGTAGY 156
       1 : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
DB     82 EHYLEDADWPLKADDODTVIIIDLNLKWLST--YDPEEPYIGRRKRKYVKGGYMGSGAGY 139
       : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : : :
QY      157 IISREAMRFELDRI 170
       ::::: :::
DB     140 VLSKEALRKPFVDAT 153

```


APPLICANT: Sherman, Amir
APPLICANT: Silva, Jeff
APPLICANT: Summers, Eric
TITLE OF INVENTION: Methods for Improving Secondary Metabolite Production in Fungi
FILE REFERENCE: 109272.147
CURRENT APPLICATION NUMBER: US/09/801,368
CURRENT FILING DATE: 2001-03-07
PRIOR APPLICATION NUMBER: US 09/487,558
PRIOR FILING DATE: 2000-01-19
PRIOR APPLICATION NUMBER: US 60/160,587
PRIOR FILING DATE: 1998-10-20
NUMBER OF SEQ ID NOS: 440
SOFTWARE: PatentIn version 3.0
SEQ ID NO 222
LENGTH: 1224
TYPE: PRF
ORGANISM: Saccharomyces cerevisiae
US-09-801-368-222

Query Match 5.4%; Score 77.5; DB 10; Length 1224;
Best Local Similarity 18.6%; Pred. No. 15;
Matches 43; Conservative 47; Mismatches 90; Indels 51; Gaps 11;

QY 17 HIAIDFGSGRIGPQFEDALMPSERLQAMAEVLSAHPYLSGVEK---ACFMSHAVL 72
DB 479 YFSSPFGQVLSSTFLD-----HKLE-----PYLGALSQYMIVECFINGCIR 520
QY 73 W-----EGALDEGLPIYAFEDDVLLGSGAQFLAEDTWLEERFDKDSAFIVRL 122
DB 521 WKIWTGTGDYDEKIDSLQKLEILSNQLIALN---LREPLLRKQIQONFALFTMLK 575
QY 123 TMEAKVIVPRDKVLYENRSPFLSEHCGTAGYIIS--REARFPLDRFAVLPPERIK 180
DB 576 DNVLEFLL--EKIITSATNDYPRINEERGAESDAVRDLRACGIELNRMALIMPESLKK 633
QY 181 VDIAMFTYFDEGAMPYOVSPALCTOE-LHYAKETLSQNSMLGS-DLEKDR 229
DB 634 I-----YPDLESV-IARIMPLSYHEKISFSEFLIIVLKSLDMKEER 676

RESULT 15
US-09-976-059-14
Sequence 14, Application US/09976059
Patent No. US20020164747A1
GENERAL INFORMATION:
APPLICANT: Farnet, Chris
APPLICANT: Zazopoulos, Emmanuel
APPLICANT: Stafia, Alfredo
TITLE OF INVENTION: Genes and Proteins for Biosynthesis of Ramoplanin
FILE REFERENCE: 3019-PCT
CURRENT APPLICATION NUMBER: US/09/976,059
CURRENT FILING DATE: 2001-10-15
NUMBER OF SEQ ID NOS: 34
SOFTWARE: PatentIn version 3.0
SEQ ID NO 14
LENGTH: 4999
TYPE: PRF
ORGANISM: Actinoplanes sp.
US-09-976-059-14

Query Match 5.4%; Score 77.5; DB 9; Length 4999;
Best Local Similarity 21.8%; Pred. No. 96;
Matches 55; Conservative 38; Mismatches 100; Indels 59; Gaps 13;

QY 5 VISLASAARRRAHI-----ADTFGS---RGIPOFFDALMPSERLQAMAEVPGISA 54
DB 2896 LVNMGGITETTVYHQDLAPADDTGSSPIGRIGP-----LSVYVLDALRPVPPGVAG 2949
QY 55 HPIYLSGVEKACFMSHAVLEQALD-----EGLPIYAVFE-----DDVLLGSGAQFLAF- 103
DB 2950 EYVYVAGRQ-----LARYLYLGRALITGRFVACFPPLPAGERMYRTGDRARWSRGRLOPAGRT 3005
QY 104 DTWLE-----ERFDKDSAFIVRLTMAKVIIVRPD-----KVLNENRSPFLSEHCG 152

DB 3006 DDQVQIRGFRIEPEGEVQAVVAHAHPEIAAAAVVREDVPGDPRLTAYVPAGP-----R 3058
QY 153 TACIYISREARFELDRF-AVLPPEKAVDLMFTYF--FDKEGMPYIQ-----VSP 202
DB 3059 TAPAAVAETVRRPFAADRLPAYMLPSAVVLDALPLTDHGKIDRRALPAQHTGAASGRAP 3118
QY 203 ALCTOELHYAKF 214
DB 3119 ATVAEEVLCNAF 3130

Search completed: December 2, 2002, 12:16:05
Job time : 8.61203 secs

THIS PAGE BLANK (USPTO)

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 12.4266 Seconds
(without alignments)
2166.126 Million cell updates/sec

Title: US-10-007-267-6
Perfect score: 1440
Sequence: 1 MGNHVISLASAERRAHAD.....ERQQAELKVVGRVILFK 280

Scoring table: BL0SUM62
Gapop 10.0 , Gapext 0.5

Searched: 283224 seqs, 96134422 residues

Total number of hits satisfying chosen parameters: 283224

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : PIR_73:*
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1284	89.2	280	A81027	lacto-N-neotetraos
2	1236	85.8	276	S70815	glycosyl transfera
3	1012	70.3	275	C81027	lacto-N-neotetraos
4	1007	69.9	268	B81971	lacto-N-neotetraos
5	1007	69.9	275	S70814	glycosyl transfera
6	971.5	67.5	279	A81971	lacto-N-neotetraos
7	500.5	34.8	266	S71025	lipopolysaccharide
8	438.5	30.5	302	A64077	lipopolysaccharide
9	315	21.9	282	F64091	lipopolysaccharide
10	233	16.2	268	A03304	lipopolysaccharide
11	121	8.4	738	T00343	lipooligosaccharid
12	120.5	8.4	332	E71916	hypothetical prote
13	108.5	7.5	267	E64175	probable lipopolys
14	104	7.2	273	B64623	hypothetical prote
15	99	6.9	273	E71890	lipopolysaccharide
16	99	6.9	394	A70469	probable lipopolys
17	98.5	6.8	721	AH3417	aspartate transam
18	97	6.7	444	F71916	lipa protein [limp
19	97	6.7	839	F85334	probable lipopolys
20	97	6.7	1446	T04528	myosin heavy chain
21	95.5	6.6	575	B83959	myosin heavy chain
22	95.5	6.6	1242	T45976	hypothetical prote
23	95	6.6	492	T20368	myosin heavy chain
24	93.5	6.5	284	E64620	lipopolysaccharide
25	92.5	6.4	404	C64597	lipopolysaccharide
26	92.5	6.4	457	T40770	colled coil protei
27	91	6.3	419	T19837	hypothetical prote
28	90	6.2	1953	T40642	probable helicase
29	89.5	6.2	266	AG2797	conserved hypothet

30	89.5	6.2	266	2	H97576	hypothetical prote
31	89.5	6.2	594	2	I49127	intracellular prot
32	88.5	6.1	256	2	T00097	hypothetical prote
33	88.5	6.1	885	1	VCVPF2	structural protein
34	88.5	6.1	2415	1	A33733	spectrin alpha cha
35	88	6.1	292	2	D71894	probable lipopolys
36	88	6.1	998	2	S47105	myosin heavy chain
37	87.5	6.1	314	2	G71974	probable keto-acid
38	87.5	6.1	429	2	A36220	transforming prote
39	87.5	6.1	662	2	G89909	transketolase [limp
40	87	6.0	468	1	D26190	phosphoglucosate d
41	87	6.0	2970	2	T08839	polyprotein - marm
42	86	6.0	291	2	AB1384	conserved hypothet
43	85.5	5.9	653	2	AB1128	transcription anti
44	85.5	5.9	1502	2	T14278	myosin-like protei
45	85.5	5.9	1529	2	A59310	unconventional myo

ALIGNMENTS

RESULT 1
A81027
lacto-N-neotetraose biosynthesis glycosyl transferase Igte NMB1926 [imported] - Neiss
C:Species: Neisseria meningitidis
C:Date: 31-Mar-2000 #sequence_revision 31-Mar-2000 #text_change 19-Jan-2001
C:Accession: A81027
R:Rettelin, H.; Saunders, N.J.; Heidelberg, J.; Jeffries, A.C.; Nelson, K.E.; Eisen,
Hickey, E.K.; Haft, D.H.; Salzberg, S.L.; White, O.; Fleischmann, R.D.; Dougherty, B.
H.; Qin, H.; Yamaheyan, J.; Gill, J.; Scarlato, V.; Maignani, V.; Pizza, M.
Science 287, 1809-1815, 2000
A:Authors: Grandi, G.; Sun, L.; Smith, H.O.; Fraser, C.M.; Moxon, E.R.; Rappunli, R.;
A:Title: Complete genome sequence of Neisseria meningitidis serogroup B strain MC58.
A:Reference number: A81000; M01D:20175755; PMID:10710307
A:Accession: A81027
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-280 <TEP>
A:Cross-references: GB:AE002541; GB:AE002098; NID:g7227175; PIDN:AA42255.1; PID:g7222
A:Experimental source: serogroup B, strain MC58
C:Genetics:
A:Gene: NMB1926
C:Superfamily: lipopolysaccharide biosynthesis-associated protein

Query Match 89.2% Score 1284; DB 2; Length 280;
Best Local Similarity 90.4%; Pred. No. 3.3e-104;
Matches 253; Conservative 9; Mismatches 18; Indels 0; Gaps 0;

QY 1 MGNHVISLASAERRAHADTFGRGIPQFPDAMPSERLQAMAEVPGLSAHPLYSG 60
DB 1 MGNHVISLASAERRAHADTFGRGIPQFPDAMPSERLQAMAEVPGLSAHPLYSG 60
QY 61 VERKACMSHAVIMEQALDEGLPYIAVEEDVILGEGAEQFLAEDTWLEERFDKSAFIVR 120
DB 61 VERKACMSHAVIMEQALDEGLPYIAVEEDVILGEGAEQFLAEDTWLEERFDKSAFIVR 120
QY 121 LETMPKAVIVRPDKVINYNSRSPFLSEHCGTAGTISRMRPFLDFAVLPERRIRA 180
DB 121 LETMPKAVIVRPDKVINYNSRSPFLSEHCGTAGTISRMRPFLDFAVLPERRIRA 180
QY 181 VDLMEFTYFDEKGMVYOVSPALCTQELHYAKFLSQNSMLSDLEKDEQGRHRRSLK 240
DB 181 VDLMEFTYFDEKGMVYOVSPALCTQELHYAKFLSQNSMLSDLEKDEQGRHRRSLK 240
QY 241 VMFIDLKRALGKFGREKKRMREROQAELKVVGRVILFK 280
DB 241 VMFIDLKRALGKFGREKKRMREROQAELKVVGRVILFK 280

RESULT 2
S70815
glycosyl transferase E (EC 2.4.-.-) - Neisseria meningitidis
C:Species: Neisseria meningitidis

RESULT 11
T00343
hypothetical protein KIAA0584 - human (fragment)
C:Species: Homo sapiens (man)
C:Date: 01-Feb-1999 #sequence_revision 01-Feb-1999 #text_change 21-Jul-2000
C:Accession: T00343
R:Nagase, T.; Ishikawa, K.; Miyajima, N.; Tanaka, A.; Kotani, H.; Nomura, N.; Ohara, O.
DNA Res. 5, 31-39, 1998
A:Title: Prediction of the coding sequences of unidentified human genes. IX. The complete
A:Reference number: Z14086; MUID:98290545; PMID:9628581
A:Accession: T00343
A:Status: preliminary; translated from GB/EMBL/DBJ
A:Molecule type: mRNA
A:Residues: 1-738 <MG>
A:Cross-references: EMBL:AB011156; NID:g3043691; PIDN:BAA2510.1; PID:g3043692
A:Experimental source: Brain
C:Genetics:
;Note: KIAA0584

Query Match	8.4%	Score 121	DR 2	Length 738
Best Local Similarity	23.0%	Pred. No. 0.019		
Matches	61	Conservative	51	Mismatches 107
				Indels 46
				Gaps 15
QY	5	VISLASAERRAHIAADTGGSRGIPQFDAL----	MPSERLQOAMVELPG----	LSAHP 56
	:	:	:	:
Db	458	MINLKRDRDRDMRLTLTLEOEIVKRIEAVDGA	KALNTSQLKALNIEMLPGPRDPYSRP	517
QY	57	YLSGVCKECPMSHAWIMQCALDEGLPYIAVFED	DVLLGEAGQPLAE----	DPWLERRFPKD 114
	:	:	:	:
Db	518	LTRG-EIGCFLSHYSVMKEVIDRELEKTLVIED	VRPEHQFRKKLTKMLMDINDAQDLWE	576
QY	115	SAFIVRETFMAKAVIVRDK-VLNTENRSFPLESE	HC-GTAGYIISREAMRFL--	DRF 170
	:	:	:	:
Db	577	LIYIGRKEMQ-----VKEPEKAVPNVAN----	LVEADYVYWTLYGYVISEGAKOLGANPF	628
QY	171	A-VLPPERIKAV-----DLMFTYFEDKEGMP	VPYQVSALCTQELHYA----	KETS----- 216
	:	:	:	:
Db	629	GKMLPVDFELPYMKHVALEYKEYESRDLKAFSA	P-LLIYPHYIGQFGYLSDTETS	687
QY	217	---ONSMUGSLE-----	KDRBQGR 233	
	:	:	:	:
Db	688	TIMDETATVATDMDRTHAUKSRKQSR	712	

RESULT 12
E71916
probable lipopolysaccharide biosynthesis protein - *Helicobacter pylori* (strain J99)
C:Species: *Helicobacter pylori*
A:Variety: strain J99
C:Date: 12-Feb-1999 #sequence_revision 12-Feb-1999 #text_change 08-Oct-1999
C:Accession: E71916
R:Alm, R.A.; Linn, L.S.L.; Molr, D.T.; King, B.L.; Brown, E.D.; Doid, P.C.; Smith, D.R.;
Iyev, C.; Gibson, R.; Merberg, D.; Mills, S.D.; Jiang, Q.; Taylor, D.E.; Voris, G.F.;
Nature 397, 176-180, 1999
A:Title: Genomic sequence comparison of two unrelated isolates of the human gastric path
A:Reference number: A71800; MUID:99120557; PMID:9923682
A:Accession: E71916
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-332 <ARN>
A:Cross-references: GB:AE001489; GB:AE001439; NID:g4155102; PIDN:AD06135.1; PID:g4155101
A:Experimental source: strain J99
C:Genetics:
C:Gene: jhp0562

```

Query Match      8.44; Score 120.5; DB 2; Length 333;
Local Similarity 22.6%; Pred. No. 0.0078;
Match 79; Conservative 47; Mismatches 98; Indels 125; Gaps 21;

QY 4 HVISLASAERRAHIAIDT-----FGSRGIPFFPDALMESER-LEQAMAEIVPGL 52
      ::::  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :
5 YIISKEQOR---LDMEKLVLENEKFGRCV-FQTFDAISPHQOEKIDLYIAQ 59
      :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :  :

```

```

OY 53 SA-----HPY---L$GVEKACEM$HAU$LM$QALDE$LP$IA$VFED$VLL$GEGA$QFLA 102
      | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 60 $LLO$D$M$W$H$Y$G$A$O$A$L$T$L$P$E$G$Y$H$Y$L$L$K$E$C$K$I$D$P$-V$I$E$D$E$T$E$H$M$Q$A$L- 117
OY 103 E$D$T$E$E$R$P$D$K$O$A$F$V$L$-----E$T$M$F$---A$K$V$P$R$D$K$V$L$Y$N$B$R$P$L$E$S$--- 148
      | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 118 E$D$C$-L$K$S$P$D$-----F$V$A$L$G$Y$G$W$Y$H$E$T$K$H$V$L$P$K$E$V$P$P$P$D$H$S$F$K$N$--P$I$E$K$K 169
OY 149 -----E$H$-----C$G$A$G$Y$I$S$R$E$A 162
Db 170 F$E$D$V$R$F$L$N$S$T$H$K$V$I$H$Y$L$L$K$O$K$S$Y$A$T$H$E$K$A$F$E$L$H$E$Y$L$T$S$V$A$S$T$A$G$Y$Y$L$P$K$G 229
OY 163 M$R$F$L$---D$R$E$A$V$L$P$P$E$R$I$K$A$V$L$D$L$M$F$T$P$D$K$E$G$M$P$Y$O$V$S$P$A$L$C$Q$E$L$H$A$K$F$L$O$N$S 219
      | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 230 A$K$T$E$A$R$E$S$F$K$I$E$P$-----V$D$M$D$N$S$A$H$D$V$A$N$L$Y$Y$P$C$V$-----S$L$E$H$S 274
OY 220 M$G$S$D$L$E$D$R$E$O$G$R$-----H$R$S$L$K$V$M$F$---D$L$K$A$L$C$F$G$R$E$K$K 258
      | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : | : : : : |
Db 275 -L$D$T$O$K$P$O$K$S$L$K$S$Y$P$L$P$O$K$S$T$F$K$N$L$F$Y$S$L$A$K$K$R$L$A$F$O$O$Y$S$ 322

```

RESULT 13
E64175
Hypothetical protein H11697 (1sg locus) - Haemophilus influenzae
C:Species: Haemophilus influenzae
C:Date: 18-Aug-1995 #sequence_revision 18-Aug-1995 #text_change 08-Oct-1999
C:Accession: E64175; S27580
R:Rifleschmann, J.D.; Adams, M.D.; White, O.; Clayton, R.A.; Kirkness, E.F.; Kerlavage,
J.; Gocayne, J.D.; Scott, J.; Shirley, R.; Liu, L.I.; Glodet, A.; Kelley, J.M.; Weidman,
D.M.; Brandon, R.C.; Fine, L.D.; Fritchman, J.L.; Fuhrmann, J.L.; Geoghagen, N.S.M.,
Science 269, 496-512, 1995
A:Authors: Gnehm, C.L.; McDonald, L.A.; Small, K.V.; Fraser, C.M.; Smith, H.O.; Ventec
A:Title: Whole-genome random sequencing and assembly of Haemophilus Influenzae Rd.
A:Reference number: A64000; MUID:95350630; PMID:7542800
A:Accession: E64175
A:Status: nucleic acid sequence not shown; translation not shown
A:Molecule type: DNA
A:Residues: 1-267 <TIGR>
A:Cross-references: GB:I42023; TIGR:H11697
A:Experimental source: strain Rd KW20
R:McLaughlin, R.; Abu Kwaik, Y.; Young, R.; Spinola, S.; Apicella, M.
submitted to the EMBL Data Library, June 1992
A:Description: Characterization and sequence of the 1sg locus from Haemophilus influenzae
A:Reference number: S27577
A:Accession: S27580
A:Molecule type: DNA
A:Residues: 11-26, 'K', 28-39, 'I', 41-55, 'S', 57-111, 'L', 113-136, 'N', 138-139, 'N', 141-143,
A:Cross-references: EMBL:M94855; NID:9148931; PIDN:AAA24981.1; PID:9148935
A:Experimental source: strain A2

[illegible]

GenCore version 5.1.3
Copyright (c) 1993 - 2002 Compugen Ltd.

OM protein - protein search, using sw model

Run on: December 2, 2002, 11:55:41 ; Search time 6.8146 Seconds
(without alignments)
1704.189 Million cell updates/sec

Title: US-10-007-267-6
Perfect score: 1440
Sequence: 1 MONHVISLASAERAHIAAD.....EORQAELEKRYGRVILFK 280

Scoring table: BIOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 112892 seqs, 41476328 residues
Total number of hits satisfying chosen parameters: 112892

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : SwissProt_40:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1440	100.0	280	1	LGTE_NEIGO
2	1284	89.2	280	1	LGTE_NEIMB
3	1036.5	72.0	279	1	LGTE_NEIGO
4	1012	70.3	275	1	LGTE_NEIMB
5	971.5	67.5	279	1	LGTE_NEIMA
6	500.5	34.8	266	1	LGTE_NEIMA
7	438.5	30.5	302	1	LGTE_NEIMA
8	331.5	23.0	263	1	LGTE_NEIMA
9	315	21.9	282	1	LGTE_NEIMA
10	108.5	7.5	257	1	LGTE_NEIMA
11	99	6.9	394	1	LGTE_NEIMA
12	95.5	6.6	933	1	LGTE_NEIMA
13	89.5	6.2	627	1	LGTE_NEIMA
14	88.5	6.1	885	1	LGTE_NEIMA
15	87.5	6.0	429	1	LGTE_NEIMA
16	87	6.0	468	1	LGTE_NEIMA
17	85	5.9	618	1	LGTE_NEIMA
18	85	5.9	762	1	LGTE_NEIMA
19	84.5	5.9	2415	1	LGTE_NEIMA
20	84	5.8	526	1	LGTE_NEIMA
21	84	5.8	512	1	LGTE_NEIMA
22	83.5	5.8	550	1	LGTE_NEIMA
23	83.5	5.8	856	1	LGTE_NEIMA
24	83.5	5.8	1178	1	LGTE_NEIMA
25	83	5.8	374	1	LGTE_NEIMA
26	83	5.8	705	1	LGTE_NEIMA
27	82.5	5.7	434	1	LGTE_NEIMA
28	82.5	5.7	881	1	LGTE_NEIMA
29	82	5.7	512	1	LGTE_NEIMA
30	82	5.7	549	1	LGTE_NEIMA
31	82	5.7	1012	1	LGTE_NEIMA
32	82	5.7	2561	1	LGTE_NEIMA
33	81.5	5.7	663	1	LGTE_NEIMA

34	81.5	5.7	2511	1	FAS_CHICK	P12276	gallus gall
35	81.5	5.7	3924	1	ANK2_HUMAN	Q01484	homo sapien
36	81	5.6	870	1	SVV_CAMJF	Q99964	campylobact
37	81	5.6	1274	1	SREB3_BACSU	O08787	bacillus su
38	80.5	5.6	714	1	YFCX_ECOLI	P77399	escherichia
39	80.5	5.6	4344	1	DYHC_EMENT	P45444	emeritocella
40	79.5	5.5	397	1	PGK_CLOPE	O8XK00	clostridium
41	79.5	5.5	779	1	VEZA_HUMAN	O9HBM0	homo sapien
42	79	5.5	286	1	SPEE_STRPN	O97RA7	streptococc
43	79	5.5	359	1	Y130_ARCNU	O30107	archaeoglob
44	79	5.5	597	1	ROP_DROME	O07327	drosophila
45	79	5.5	1498	1	Y1A9_CLOAB	Q04351	clostridium

ALIGNMENTS

RESULT 1
ID LGTE_NEIGO STANDARD: PRT; 280 AA.
AC Q50950:

DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lacto-N-neotetraose biosynthesis glycosyl transferase lgte.
GN LGTE.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=FB2;
RX MEDLINE=95053752; PubMed=7964493;
RA Gotschlich E.C.;
RT "Genetic locus for the biosynthesis of the variable portion of
RT Neisseria gonorrhoeae lipooligosaccharide";
RL J. Exp. Med. 180:2181-2190(1994).
CC -1- FUNCTION: ADDS THE FIRST GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN
CC IN LOS.
CC -1- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS
CC STRUCTURE.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb.ch/announce/>
CC or send an email to license@isb.ch).

DR EMBL: U14554; AAA68013.1; -;
DR InterPro: IPR002654; Grf_25;
DR Pfam: PF01755; Glyco_Transf_25; 1;
KW Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase.
SQ SEQUENCE 280 AA; 32420 MW; F043597BA3040407 CRC64;

Query Match 100.0%; Score 1440; DB 1; Length 280;
Best Local Similarity 100.0%; Pred. No. 1.2e-113; Indels 0; Gaps 0;
Matches 280; Conservative 0; Mismatches 0;

QY	1	MONHVISLASAERAHIAADTFGSGRIFPOFPDAMPSERLQAAELVPGLSAHPIYSG	60
DB	1	MONHVISLASAERAHIAADTFGSGRIFPOFPDAMPSERLQAAELVPGLSAHPIYSG	60
QY	61	VERACFMSHAVLMEQALDEGLPYIAVFEDDVLGAGABOFLAEDTWLEBEREDKDSAFIVR	120
DB	61	VERACFMSHAVLMEQALDEGLPYIAVFEDDVLGAGABOFLAEDTWLEBEREDKDSAFIVR	120
QY	121	LETMAKIVIPDKVLNENSGFLLSEHCGTACIYISREAMRFIDRFVAVLPERIKA	180
DB	121	LETMAKIVIPDKVLNENSGFLLSEHCGTACIYISREAMRFIDRFVAVLPERIKA	180

```

OY 161 VDLMMFTYFDEKGMYPYOVSPALCTOELHYAKFLTSQNSMLGSDLEKDEQRRRRRSIK 240
    |||||||
DB 181 VDLMMFTYFDEKGMYPYOVSPALCTOELHYAKFLTSQNSMLGSDLEKDEQRRRRRSIK 240
OY 241 VMEFLKRALGKFGREKKRMROROAELKYGRRVILFK 280
    |||||||
DB 241 VMEFLKRALGKFGREKKRMROROAELKYGRRVILFK 280

RESULT 2
LGTE_NEIMB STANDARD: PRT: 280 AA.
ID LGTE_NEIMB STANDARD: PRT: 280 AA.
AC 05117;
DT 01-NOV-1997 (Rel. 35, Created)
DT 16-OCT-2001 (Rel. 40, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lacto-N-neotetraose biosynthesis glycosyl transferase lgte.
GN LGTE OR NMB1926.
OS Neisseria meningitidis (serogroup B).
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=491;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=96414473; PubMed=8617494;
RA Jennings M.P., Hood D., Peak I.R.A., Virji M., Moxon E.R.;
RT "Molecular analysis of a locus for the biosynthesis and phase-variable
RT expression of the lacto-N-neotetraose terminal lipopolysaccharide
RT structure in Neisseria meningitidis.";
RL Mol. Microbiol. 18:729-740(1995).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=MC58 / Serogroup B;
RX MEDLINE=20175755; PubMed=10710307;
RA Tettelin H., Saunders N.J., Heidelberg J., Jeffries A.C., Nelson K.E.,
RA Eisen J.A., Ketchum K.A., Hood D.W., Peden J.F., Dodson R.J.,
RA Nelson W.C., Gwin M.L., DeBoy R., Peterson J.D., Hickey E.K.,
RA Haft D.H., Salzberg S.L., White O., Fleischmann R.D., Dougherty B.A.,
RA Mason T., Ciecko A., Parksey D.S., Blair E., Clifton H., Clark E.B.,
RA Cotton M.D., Ulterbach T.R., Khouvri H., Qin H., Vamathevan J.,
RA Gill J., Scarlato V., Maignani V., Pizza M., Grandi G., Sun L.,
RA Smith H.O., Fraser C.M., Moxon E.R., Rappunoi R., Venter J.C.;
RT "Complete genome sequence of Neisseria meningitidis serogroup B strain
RT MC58.";
RL Science 287:1809-1815(2000).
CC -1- FUNCTION: ADDS THE FIRST GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN
CC IN LOS.
CC -1- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS
CC STRUCTURE.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U25839; AAC4086.1; -
CC DR EMBL: AE002541; AAF42255.1; -
CC DR TIGR: NMB1926; -
CC DR InterPro: IPR002654; GT_25.
CC Pfam: PF01755; Glyco_transf_25; 1.
CC Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;
CC Complete proteome.
CC KW COMPLETE 16
CC FT CONFLICT 16 A -> G (IN REF. 1).
CC FT CONFLICT 89 MISSING (IN REF. 1).
CC FT CONFLICT 176 EMI -> RVD (IN REF. 1).
CC FT CONFLICT 178 EMI -> RVD (IN REF. 1).
CC SEQUENCE 280 AA; 32790 MW; 826827E942BA5842 CRC64;

Query Match 89.2%; Score 1284; DB 1; Length 280;
Best Local Similarity 90.4%; Pred. No. 1.5e-100;

```

```

Matches 253; Conservative 9; Mismatches 18; Indels 0; Gaps 0;
OY 1 MONHYISLASAERRAHIAADTFGSGRIFPOFDALMPSERLQAAELVPGLSAHPYLSG 60
    |||||||
DB 1 MONHYISLASAERRAHIAADTFGSGRIFPOFDALMPSERLQAAELVPGLSAHPYLSG 60
OY 61 VEKACFMSHAVIMEQALDEGLPIYAVFEDDVLGCGAEOFLAEDTWLEBRDKDSAFIVR 120
    |||||||
DB 61 VEKACFMSHAVIMEQALDEGLPIYAVFEDDVLGCGAEOFLAEDTWLEBRDKDSAFIVR 120
OY 121 LETMFAKVIYRPDKVLNENRSFPLESEHOGTAGYIISREAMRFELRPATVLPERRIKA 180
    |||||||
DB 121 LETMFAKVIYRPDKVLNENRSFPLESEHOGTAGYIISREAMRFELRPATVLPERRIKA 180
OY 181 VDLMMFTYFDEKGMYPYOVSPALCTOELHYAKFLTSQNSMLGSDLEKDEQRRRRRSIK 240
    |||||||
DB 181 VDLMMFTYFDEKGMYPYOVSPALCTOELHYAKFLTSQNSMLGSDLEKDEQRRRRRSIK 240
OY 241 VMEFLKRALGKFGREKKRMROROAELKYGRRVILFK 280
    |||||||
DB 241 VMEFLKRALGKFGREKKRMROROAELKYGRRVILFK 280

RESULT 3
LGTE_NEIGO STANDARD: PRT: 279 AA.
ID LGTE_NEIGO STANDARD: PRT: 279 AA.
AC 050947;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lacto-N-neotetraose biosynthesis glycosyl transferase lgte.
GN LGTB.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=F62;
RX MEDLINE=95053752; PubMed=7964493;
RA Gotschlich E.C.;
RT "Genetic locus for the biosynthesis of the variable portion of
RT Neisseria gonorrhoeae lipooligosaccharide.";
RL J. Exp. Med. 180:2181-2190(1994).
CC -1- FUNCTION: ADDS THE SECOND GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN
CC IN LOS.
CC -1- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS
CC STRUCTURE.
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
CC EMBL: U14554; AAA68010.1; -
CC DR EMBL: AE002654; GT_25.
CC DR InterPro: IPR002654; GT_25.
CC Pfam: PF01755; Glyco_transf_25; 1.
CC Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase.
CC KW LIPOPOLYSACCHARIDE BIOSYNTHESIS;
CC SEQUENCE 279 AA; 31776 MW; C2219D1A6119D622 CRC64;

Query Match 72.0%; Score 1036.5; DB 1; Length 279;
Best Local Similarity 71.6%; Pred. No. 8.1e-80;
Matches 209; Conservative 24; Mismatches 34; Indels 25; Gaps 3;
OY 1 MONHYISLASAERRAHIAADTFGSGRIFPOFDALMPSERLQAAELVPGLSAHPYLSG 60
    |||||||
DB 1 MONHYISLASAERRAHIAADTFGSGRIFPOFDALMPSERLQAAELVPGLSAHPYLSG 60
OY 61 VEKACFMSHAVIMEQALDEGLPIYAVFEDDVLGCGAEOFLAEDTWLEBRDKDSAFIVR 120
    |||||||
DB 61 VEKACFMSHAVIMEQALDEGLPIYAVFEDDVLGCGAEOFLAEDTWLEBRDKDSAFIVR 120

```

SQ SEQUENCE 275 AA; 31578 MW; E871305E2F6CF70F CRC64;
 Query Match 70.3%; Score 1012; DB 1; Length 275;
 Best Local Similarity 75.18; Pred. No. 9e-78;
 Matches 202; Conservative 20; Mismatches 43; Indels 4; Gaps 1;
 QY 1 MGNVHSLASAERRAHNADTFGSGIPGFPPDAMPSEERLEQAAEVLPGLSAHPLYSG 60
 Db 1 MGNVHSLASAERRAHNADTFGSGIPGFPPDAMPSEERLEQAAEVLPGLSAHPLYSG 60
 QY 61 VEKACFMSHAVLMEQALDEGLPYIAVEFDVYLLGGAEOFLAEDTLEERPKDSAFIVR 120
 Db 61 VEKACFMSHAVLMEQALDEGLPYIAVEFDVYLLGGAEOFLAEDTLEERPKDSAFIVR 120
 QY 121 LETFAKIVAPDVIENRSEPLTSEHOGTGAYITISREAMREFLIRPVLPERIKA 180
 Db 121 LETFAKIVAPDVIENRSEPLTSEHOGTGAYITISREAMREFLIRPVLPERIKA 180
 QY 181 VDLMEFTYFPDEKGNPVYQVSPALCTOELHYAKFLSQNSMLGSDLEKD----REGRRHR 236
 Db 181 VDLMEFTYFPDEKGNPVYQVSPALCTOELHYAKFLSQNSMLGSDLEKD----REGRRHR 236
 QY 237 RSLKVMFDLKRALGKFGREKKKKRMRQRO 265
 Db 241 PANTFKNRLIRALTKISRERKRQRQRO 269
 RESULT 5
 LGTB_NEITMA STANDARD: PRT; 279 AA.
 AC PS7033;
 DT 16-OCT-2001 (Rel. 40, Created)
 DT 16-OCT-2001 (Rel. 40, Last sequence update)
 DT 15-JUN-2002 (Rel. 41, Last annotation update)
 DE Lacto-N-neotetraose blosynthesis glycosyl transferase lgTB.
 GN LGTB OR NMA0525.
 OS Neisseria meningitidis (serogroup A).
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_TaxID=65699;
 RN (1)
 RP SEQUENCE FROM N.A.
 RC STRAIN=22491 / Serogroup A / Serotype 4a;
 RX MEDLINE=20222556; PubMed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Devlin K., Feltham T., Hamlin N., Holroyd S.,
 RA Jagsels K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.;
 RA "Complete DNA sequence of a serogroup A strain of Neisseria
 RT meningitidis 22491.";
 RL Nature 404:502-506(2000).
 CC -!- FUNCTION: ADDS THE SECOND GALACTOSE TO THE LACTO-N-TETRAOSE CHAIN
 CC IN LOS.
 CC -!- PATHWAY: BIOSYNTHESIS OF THE TERMINAL LACTO-N-NEOTETRAOSE LPS
 CC STRUCTURE.
 CC -!- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (see <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC DR EMBL; AL162753; CAB83817.1; -
 CC InterPro: IPR002654; GT_25.
 CC Pfam: PF01755; Glyco_transf_25; 1.
 CC Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase;
 CC Complete proteome.
 CC SEQUENCE 279 AA; 31903 MW; 8703B56513A0D347 CRC64;

CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL, U32736; AAC22208.1; -
DR EMBL, X56903; CAA40221.1; -
DR EMBL, L19441; AAA65534.1; -
DR PIR, S15282; S15282.
DR TIGR, H10550; -
DR InterPro, IPR002654; GT_25.
DR Pfam, PF01755; Glyco_transf_25; 1.
KM Transferase; Glycosyltransferase; Virulence; Complete proteome.
FT VARIANT 29 29 K -> N (IN STRAIN DL42 AND RM 7004).
FT VARIANT 63 70 MISSING (IN STRAIN RM 7004).
FT VARIANT 67 70 MISSING (IN STRAIN DL42).
FT VARIANT 106 106 F -> L (IN STRAINS DL42 AND RM 7004).
FT VARIANT 151 151 Q -> R (IN STRAINS DL42 AND RM 7004).
FT VARIANT 256 256 S -> P (IN STRAIN RM 7004).
FT VARIANT 274 274 N -> D (IN STRAIN RM 7004).
SQ SEQUENCE 302 AA; 35490 MW; BFC204F9B3372D2C CRC64;

Query Match 30.5%; Score 438.5; DB 1; Length 302;
Best Local Similarity 33.0%; Pred. No. 1.2e-29;
Matches 102; Conservative 68; Mismatches 98; Indels 41; Gaps 7;

QY 1 MONHYISLASAERAHADTFGSRGIPQFPDAL-----MPSERL 41
DB 4 IENYISMENATEERKHHITKOFESKLSFFNATYQISINOSINOSINOSINOSTI 63
QY 42 EQAMAE--LVPGLSAHPYLSGVERACFMSHAVLMEQALDEGLPIYAVEEDVLLGEGAE 98
DB 64 NOSINOSINILNIEESRLITLTKGKCLISHFLMNKCVNEMNEVYLKTEDEVDYILGENAE 123
QY 99 QFLAEDTWLEERFDKDSATIVLETWFAKV-IVRPDKVLYANTENRSPPLLESEHCAGTII 157
DB 124 VFLNEMWETKTFDFENDIFIRLETFLOPKLEKOTKIPPFNSRNDILKSTHWGTAGYI 183
QY 158 ISREARFELDFPFAVLPERRIKAVDLMTYFEDKGMVYOVSPALCTOELHYAKFLSQ 217
DB 184 ISOGAAKYIETLKNIPSEIYAVDELINKNLVDNNTYVQLNPICIQEOLQAND---S 240
QY 218 NSMLGSDLEKDEQGRHRH-----RSLKWFDLKRALGKFGREKKRMRQAELEK 270
DB 241 KSVLTSGLEKEREKRSKIRKKTLLKQRLTRIKENITRAL-----NRKKMKEDQRIEMQ- 294
QY 271 VVGRVILF 279
DB 295 --GKEIVRF 301

RESULT 8

LPSA_PASHA
ID LPSA_PASHA STANDARD: PRT: 263 AA.
AC 005770;
DT 01-JUN-1994 (Rel. 29, Created)
DT 01-FEB-1996 (Rel. 33, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Lpsa protein.
GN LPSA.
OS Pasteurella haemolytica.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Mannheimia.
OX NCBI_TaxID=75985;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Serotype A1;
RX MEDLINE=95301198; PubMed=7781993;
RA Potter M.D., Lo R.Y.C.;
RT "Cloning and in lipopolysaccharide of a gene from Pasteurella haemolytica
RL FEMS Microbiol. Lett. 129:75-81(1995).
RN [2]

RP SEQUENCE OF 189-263 FROM N.A.

RC STRAIN-Serotype A1;

RX MEDLINE=91358346; PubMed=1885539;

RA Abdullah K.M., Lo R.Y.C., Mellors A.;

RT "Cloning, nucleotide sequence, and expression of the Pasteurella

haemolytica A1 glycoproteinase gene.";

RL J. Bacteriol. 173:5597-5603(1991).

CC -1- FUNCTION: INVOLVED IN THE BIOSYNTHESIS OF A MOIETY ON THE CORE

OF THE LIPOLYSACCHARIDE MOLECULE.

CC -1- SUBCELLULAR LOCATION: CYTOPLASMIC. ASSEMBLED AT THE INNER SURFACE

OF THE CYTOPLASMIC MEMBRANE (POTENTIAL).

CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.

CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
CC or send an email to license@isb-sib.ch).

DR EMBL, U15958; AAA80283.1; -

DR InterPro, IPR002654; GT_25.

DR Pfam, PF01755; Glyco_transf_25; 1.

KM Lipopolysaccharide biosynthesis; Transferase; Glycosyltransferase.

SQ SEQUENCE 263 AA; 30871 MW; B07C39EC1DA603A7 CRC64;

Query Match 23.0%; Score 331.5; DB 1; Length 263;
Best Local Similarity 35.2%; Pred. No. 9e-21;
Matches 100; Conservative 41; Mismatches 118; Indels 25; Gaps 9;

QY 1 MONHYISLASAERAHADTFGSRGIPQFPDALMPSEERLQAMAEVPGLSAHPYLSG 60
DB 1 MNHYISLSAERKHHITKOFESKLSFFNATYQISINOSINOSINOSINOSTI 58
QY 61 VEKACFMSHAVLMEQALDEGLPIYAVEEDVLLGGAQFLAEDTWLEERFDKDSAFIVR 120
DB 59 GEIACALSHIALMHLAKQNLIDYICIFEDDIYLGNNAPFL-----LKTNYIPEMTHIVK 112
QY 121 LETMAKVYVRDQVLYN-ENRSPFLSEHCAGTIIISREARFELDFRFAVLPERRIK 179
DB 113 LETLPDRINRNRKTEKYLINRLFKLNSRHVGTAGYIILTNGAFLN---ILKTLNIP 169
QY 180 AVDLMAFTYFPEKEMPVYQVSPALCTOELHYAKFLSONMSGSDLEKDEQGRHRSL 239
DB 170 IDDLIFDEYLKIKE-YKVQNSPALCYQDF-----ILNSTNKKSSIQDDRALRCNNEDI 224
QY 240 KVFMDLKR--LGRFGREKKRMRQAELEKYGRVILF 280
DB 225 KNOAKLKRPNYFLTKIQKELYPFKQLKQ---KVF-TYITFK 263

RESULT 9

Y765_HAETN
ID Y765_HAETN STANDARD: PRT: 282 AA.
AC 057125; 005033;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE Hypothetical glycosyl transferase H10765 (EC 2.-.-.-).
GN H10765.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=9530630; PubMed=7542800;
RA Fieschmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Giolek A., Kelley J.M.,

```

RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uteerback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.:
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RT Science 269:496-512(1995).
CC -1- SIMILARITY: BELONGS TO THE GLYCOSYLTRANSFERASE FAMILY 25.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC DR EMBL; U32760; AAC22423.1; -.
CC DR TIGR; H10765; -.
CC DR InterPro: IPR002654; GT_25.
CC DR Pfam; PF01755; Glyco_transf_25; 1.
CC KW Hypothetical protein; Transferase; Glycosyltransferase;
CC Complete proteome.
CC SQ SEQUENCE 282 AA; 32481 MW; D962E402AA207A18 CRC64;

Query Match 21.9%; Score 315; DB 1; Length 282;
Best Local Similarity 30.9%; Pred. No. 2,4e-19;
Matches 82; Conservative 54; Mismatches 105; Indels 24; Gaps 9;

QY 3 NHVLSLSAERRAHIAIDTFGSGRIGFQFPDAMPSERLQAMAEIVPGLSAHP--YLSG 60
DB 34 NYVISLTFEQRKRKHITBEEGKQNIPEFPDALT-P-DIIEFTAKKNITLDRSPKAKLSD 92
QY 61 VERACRSHAVLMEQALDEGLPIAVFEDVLLGEGAEQFLADDTLLEEFKDSAFYR 120
DB 93 GEICGALSHVIVMDLLENLNYINIFEDDIHGENAKLELDID-----YISDDIHVLK 146
QY 121 LEMFAFVIYRPDKLVNENRSFPLESEHGTAGYIISREARFPLDRFAVLPERRIKA 180
DB 147 LEAN-GKMFEPKQKSVKCDNVPMTYKQ-SGCAGYITVTKGAKYILLEKKNPLD--VA 202
QY 181 VDIMFTYFDEKGMYPYQSPALCTQELHYAKFLSONSMGLSDE--KDREGRRHRHS 238
DB 203 VDSLVEFEDFHFHFDYKIVQLSPGICVODF---VLHDPNFESSLOGRDRVANGNQKSS 258
QY 239 LKYMFDLKRALGKREKRRKRMERQ 263
DB 259 I-----LEKIKNEFGVKIKMFGKQ 278

RESULT 10
LSG4_HAEIN STANDARD; PRT; 257 AA.
ID LSG4_HAEIN
AC P71398; Q48213;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE LSG locus putative protein 4.
GN H11697.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OC Haemophilus.
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-A2;
RA McLaughlin R., Abu Kwaik Y., Young R., Spinoia S., Apicella M.;
RT "Characterization and sequence of the lsg locus from Haemophilus
RT influenzae."
RL Submitted (JUN-1992) to the EMBL/GenBank/DBJ databases.
RN [2]
RP SEQUENCE FROM N.A.
```

```

RC STRAIN-Rd / KW20 / ATCC 51907;
RX MEDLINE=95350630; PubMed=7542800;
RA Fleischmann R.D., Adams M.D., White O., Clayton R.A., Kirkness E.F.,
RA Kerlavage A.R., Bult C.J., Tomb J.-F., Dougherty B.A., Merrick J.M.,
RA McKenney K., Sutton G., Fitzhugh W., Fields C.A., Gocayne J.D.,
RA Scott J.D., Shirley R., Liu L.-I., Glodek A., Kelley J.M.,
RA Weidman J.F., Phillips C.A., Spriggs T., Hedblom E., Cotton M.D.,
RA Uteerback T.R., Hanna M.C., Nguyen D.T., Saudek D.M., Brandon R.C.,
RA Fine L.D., Fritchman J.L., Fuhrmann J.L., Geoghegan N.S.M.,
RA Gnehm C.L., McDonald L.A., Small K.V., Fraser C.M., Smith H.O.,
RA Venter J.C.:
RA "Whole-genome random sequencing and assembly of Haemophilus influenzae
RT Rd."
RT Science 269:496-512(1995).
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement. (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch.)
CC -----
CC DR EMBL; U32842; AAC23343.1; -.
CC DR EMBL; M94855; AAA24981.1; -.
CC DR TIGR; H11697; -.
CC DR InterPro: IPR002654; GT_25.
CC DR Pfam; PF01755; Glyco_transf_25; 1.
CC Complete proteome.
KW FT CONFLICT 17
KW FT CONFLICT 30
KW FT CONFLICT 46
KW FT CONFLICT 102
KW FT CONFLICT 127
KW FT CONFLICT 127
KW FT CONFLICT 130
KW FT CONFLICT 134
KW FT CONFLICT 153
KW FT CONFLICT 214
KW FT CONFLICT 214
KW SQ SEQUENCE 257 AA; 29864 MW; D1ECDE47763EE059 CRC64;

Query Match 7.5%; Score 108.5; DB 1; Length 257;
Best Local Similarity 22.1%; Pred. No. 0.043;
Matches 62; Conservative 51; Mismatches 110; Indels 57; Gaps 13;

QY 1 MONHVLSLSAERRAHIAIDTFCSR--GIQFQFPDAMPSERLQAMAEIVPGLSAHP 56
DB 2 LKYYLISLDKDIOR-----ELPFSQKNTEDFOYFSINIMQKMDLAIIFINIQKAAH- 56
QY 57 YLSGVEK---ACPMASHAVLMEQALDEG---LPYIAVFEDVLLGEGAEQFLAEDTWLEE 109
DB 57 YFRNVTKGEIGCTLSHLSVYQKIVEENDIAEDSYALVCEDDALFHSDFQGNLT--ALLSE 114
QY 110 RFDKDSAFI-----VLEIMFAKYIVDPKVLNENRSFPLESEHCGTNG 155
DB 115 KLEAEITLLGQSKINDFNDELEINPTTFSEFCKKRTGDV---NYAFP-YKSYFAGTVG 169
QY 156 YIISREARFPLDRFAVLPERRIKAVDLMFMYFEPKDEGMYPYQSPALCTQELHYAKFL 215
DB 170 YLTKSAARFEIQIISQNKPFWL-ADDFLLEQNEINIRKKVVRPL-----IV 215
QY 216 SONSMGLSDEKDRQ-----GRHRRLSLKVMFDLKRAL 249
DB 216 IENPVLSINLESYKSLSNLKLKMKYPLKRTFAIKKWL 255

RESULT 11
AAT_AOUAE STANDARD; PRT; 394 AA.
ID AAT_AOUAE
AC O67781;
DT 30-MAY-2000 (Rel. 39, Created)
DT 30-MAY-2000 (Rel. 39, Last sequence update)
DT 16-OCT-2001 (Rel. 40, Last annotation update)
DE Aspartate aminotransferase (EC 2.6.1.1) (Transaminase A) (AspAT).
```

```

GN  ASPC OR AQ_1969.
OS  Aquifex aeolicus.
OC  Bacteria; Aquificae; Aquificae (class); Aquificales; Aquificaceae;
OC  Aquifex.
OX  NCBI_TaxID=63363;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  STRAIN-VF5;
RX  MEDLINE=98196666; PubMed=9537320;
RA  Deckert G., Warren P.V., Gaasterland T., Young W.G., Lenox A.L.,
RA  Graham D.E., Overbeek R., Sneed M.A., Keller M., Anjay M., Huber R.,
RA  Feldman R.A., Short J.M., Olson G.J., Swanson R.V.;
RT  "The complete genome of the hyperthermophilic bacterium Aquifex
RT  aeolicus";
RL  Nature 392:353-358(1998).
CC  -!- CATALYTIC ACTIVITY: L-aspartate + 2-oxoglutarate -> oxaloacetate +
CC  L-glutamate.
CC  -!- COFACTOR: PYRIDOXAL PHOSPHATE (BY SIMILARITY).
CC  -!- SIMILARITY: BELONGS TO CLASS-1 OF PYRIDOXAL-PHOSPHATE-DEPENDENT
CC  AMINOTRANSFERASES.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL, AE000766; AAC07746.1; -.
DR  HSPS; Q56232; 1BK6.
DR  InterPro: IPR001176; ACC_synthase.
DR  InterPro: IPR004839; Aminotransf1/2.
DR  InterPro: IPR004838; NHTransf_1.
DR  Pfam: PF00155; aminotran_1.2; 1.
DR  PRINTS: PR00753; ACCSYNTHASE.
DR  PROSITE: PS00105; AA_TRANSFER_CLASS_1; FALSE_NEG.
KW  transferase; aminotransferase; pyridoxal phosphate; complete proteome.
FT  BINDING 239 239 PYRIDOXAL PHOSPHATE (BY SIMILARITY).
SQ  SEQUENCE 394 AA; 43777 MW; E570BAFD080C56E1 CRC64;

Query Match 6.9%; Score 99; DB 1; Length 394;
Best Local Similarity 22.8%; Pred. No. 0.45;
Matches 75; Conservative 38; Mismatches 126; Indels 90; Gaps 17;

QY 9 ASAEERRAHIAIDFG-----SRGIP---IQFPA 34
DB 22 AKAKELRAKGVIVIGGAGEPDPDPDFKACIRALREGKTKYVPSAGIPELREAIATK 81
QY 35 LMPSERLQAMAEIVPGLSAHPLYSGVERAKCFMSHAVLMEQALDEG-----LPYIAVE 88
DB 82 LAKENKVEKPEIIVS-----AGAKNVLF-----IFMALIDBEDVLLSPYVATVP 130
QY 89 DVLVLEGAEQFLAEDTWLEERPDKDSAFIVRLTWFARVIVRPDK-VIN-----Y 138
DB 131 EDIRFGGVV-----VEVPLKKEKGFQSLSEDEKVEKTEKTAIVINSPNPGAVY 182
QY 139 EKRSPFLSEHGCTAG-YIISREARFPL--DRFVLEPERIKAVDLMPFY--FDKE 193
DB 183 EEEELKRTI-AEFCVERGIFITISDECEYFVYDAKFVSPASFSDEKNTTFVNAASKSY 241
QY 194 GMPVYQSPALCTOELHYAKFL--SONSMGSLDEKREGRRHRRLSKLWPFDKRLGK- 251
DB 242 SMTGMAIGVYACPEE--YAKVIASLNSQSVNTTFAQIG-----ALEALKPKNSKD 291
QY 252 FGEKKKKMERQQA---ELEKVGRRVI 277
DB 292 FVEMERNAFERRDRTAVELSLKIPGMDV 320

RESULT 12
GLND_RHILIO STANDARD; PRT; 933 AA.
ID GLND_RHILIO

```

```

AC Q98C27;
DT 15-JUN-2002 (Rel. 41, Created)
DT 15-JUN-2002 (Rel. 41, Last sequence update)
DT 15-JUN-2002 (Rel. 41, Last annotation update)
DE [Protein-PII] uridylyltransferase (EC 2.7.7.59) (PII uridylyl-
DE transferase) (uridylyl removing enzyme) (UTase).
GN Rhizobium loti (Mesorhizobium loti).
OS Rhizobium loti (Mesorhizobium loti).
OC Bacteria; Proteobacteria; alpha subdivision; Rhizobiaceae group;
OC Phyllobacteriaceae; Mesorhizobium.
OX NCBI_TaxID=381;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-MAFF303099;
RX MEDLINE=21082930; PubMed=11214968;
RA Kaneo T., Nakamura Y., Sato S., Asamizu E., Kato T., Sasamoto S.,
RA Watanabe A., Idehara K., Ishikawa A., Kawashima K., Kimura T.,
RA Kishida Y., Kiyokawa C., Kohara M., Matsumoto M., Matsuno A.,
RA Mochizuki Y., Nakayama S., Nakazaki N., Shimpo S., Sugimoto M.,
RA Takeuchi C., Yamada M., Tabata S.;
RT "Complete genome structure of the nitrogen-fixing symbiotic bacterium
RT Mesorhizobium loti";
RL DNA Res. 7:331-338(2000).
CC -!- FUNCTION: Modifies, by uridylylation or deuridylylation the PII
CC (glbN) regulatory protein.
CC -!- CATALYTIC ACTIVITY: UTP + [protein-PII] = diphosphate + uridylyl-
CC [protein-PII].
CC -!- SIMILARITY: BELONGS TO THE GLND FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR  EMBL, AP003006; BAB51794.1; -.
DR  InterPro: IPR002912; ACT.
DR  InterPro: IPR002819; HD.
DR  InterPro: IPR003607; MR_Plpase_HDC.
DR  InterPro: IPR002934; NTP_transf.
DR  Pfam: PF01842; ACT; 2.
DR  Pfam: PF01966; HD; 1.
DR  Pfam: PF01909; NTP_transf_2; 1.
DR  SMART: SM00471; HDC; 1.
KW  transferase; Nucleotidyltransferase; Nitrogen fixation;
KW complete proteome.
SQ  SEQUENCE 933 AA; 105279 MW; 79BF065DAC796FP8 CRC64;

Query Match 6.6%; Score 95.5; DB 1; Length 933;
Best Local Similarity 22.1%; Pred. No. 2.6;
Matches 51; Conservative 31; Mismatches 72; Indels 77; Gaps 11;

QY 4 HVISLASAERRAHIAIDFGSRGIFQFPDALMPSERLQAMAEIVPGLSAHPLYSGVEK 63
DB 293 HHMFLTKRAEERLH-----FD-----IQRLEAERL-GYTHRPGLSAVNR 330
QY 64 ACFMSHAVLMEQALDEGLPIYAFVEDDVLVIGGAGDFOFLAEDTWLEERPDKDSAFIVRLFT 123
DB 331 --FMKH-----YFLVARD--VG-----DLTRIFCAALEE 355
QY 124 MEAKVIYRPDKV-LNVENSFPLSEHGCTAGYIISREARFPLDRFVLEPERIKAVD 182
DB 356 EQAKHVPGFNRILFTFQRRKRL-----AGTSDFYVDNRHINADQVERDP-----VN 405
QY 183 LKMTFYFDEKGMVYQSPALCTOELHYAKFLQNSMGLSLEKDRREGR 233
DB 406 LRLRFWADKHGLEHFRPDALKLITSL-----GLVYKSLRDEEANKR 447

RESULT 13
SADI_MOUSE

```



```

ID  SA01.MOUSE      STANDARD;          PRT;      627 AA.
AC  060710:
DT  01-NOV-1997 (Rel. 35, Created)
DT  15-JUN-2002 (Rel. 41, Last sequence update)
DT  15-JUN-2002 (Rel. 41, Last annotation update)
DE  SAM domain and HD domain-containing protein 1 (Interferon-gamma
DE  inducible protein Mgl1).
CN  SAMHD1 OR MG11.
OS  Mus musculus (Mouse).
OC  Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC  Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Mus.
OX  NCBI_Taxid=10090;
RN  [1]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Macrophage;
RX  MEDLINE=95190404; PubMed=7884320;
RT  Lafuze W.P., Brown D., Castle L., Zwilling B.S.;
RT  "Cloning and characterization of a novel cDNA that is IFN-gamma-
RT  induced in mouse peritoneal macrophages and encodes a putative GTP-
RT  binding protein.";
RL  J. Leukoc. Biol. 57:477-483(1995).
RN  [2]
RP  SEQUENCE FROM N.A.
RC  TISSUE=Breast;
RA  Strausberg R.;
RL  Submitted (AUG-2001) to the EMBL/GenBank/DBJ databases.
CC  -1- INDUCTION: By Interferon gamma.
CC  -1- SIMILARITY: CONTAINS 1 SAM DOMAIN.
CC  -1- SIMILARITY: CONTAINS 1 HD DOMAIN.
CC  -1- CAUTION: Ref.1 sequence differs from that shown due to a
CC  frameshift in position 576.
CC  -----
CC  This SWISS-PROT entry is copyright. It is produced through a collaboration
CC  between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC  the European Bioinformatics Institute. There are no restrictions on its
CC  use by non-profit institutions as long as its content is in no way
CC  modified and this statement is not removed. Usage by and for commercial
CC  entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC  or send an email to license@isb-sib.ch).
CC  -----
DR  EMBL; U15635; AAA6219.1; ALT_FRAME.
DR  EMBL; BC012721; AAH12721.1; -.
DR  MGD; MGI:1927468; Samhd1.
DR  InterPro; IPR002819; HD.
DR  InterPro; IPR003607; ME_Pplase_HDC.
DR  InterPro; IPR001660; SAM.
DR  Pfam; PF00536; SAM; 1.
DR  Pfam; PF01966; HD; 1.
DR  SMART; SM00471; HDC; 1.
DR  SMART; SM00454; SAM; 1.
DR  PROSITE; PS50105; SAM_DOMAIN; 1.
DR  Interferon Induction.
FW  DOMAIN 46 111 SAM.
FW  DOMAIN 165 320 HD.
FT  FT  DOMAIN 165 320 HD.
SQ  SEQUENCE 627 AA; 72650 MW; C68BB853C3F4B17C CRC64;

Query Match 6.2%; Score 89.5; DB 1; Length 627;
Best Local Similarity 22.4%; Pred. No. 5.1;
Matches 54; Conservative 34; Mismatches 76; Indels 77; Gaps 13;

OY  57 YLSGVKRCQFMSHANVLMDQALDEGLPIYAVFEDVL-----LGEAGQFLAEDTW 106
OY  174 YLAG-----CLV-----RALAEKQPELOISERDILCYOAGLCHDLGHPFSNHF---- 218
OY  107 LEERFDKSAFIVRLQETWFAKVIYPPDVVLNNENRSPFLSEHCGTAGIYSRRAMRF 166
OY  219 -----DGRFLPR-----ARPEKKWKHEGSIEMF-----HLVNSNELKLV 254
OY  167 LDRFAVLPERIKAVDLMFTYFFDK-----GNPVQVSPAL-----CTOELHAKFLS 216
OY  255 MKNVGLVPEEDIT-----FIKEDQIMGPPIIPVADSLMPYGRATKSFYLE--IV 302
OY  217 ONSMLGSDLEK-DREOGRHRRSLKVMFDLKRALKGFGREKKKKRMRORQAEELEKVGRR 275

```

```

Db 303 SNKRAGCIDVDKMDYFARBCDHGIGNNDYKRFI-KFAR-----ICEVEYKAKEDKIYIK 357
Oy 276 V 276
Db 358 V 358

RESULT 14
VNCS_BMDNV STANDARD: PRG: 885 AA.
AC P05842:
DT 01-NOV-1988 (Rel. 09, Created)
DT 01-NOV-1988 (Rel. 09, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE Putative noncapsid protein NS-1 (Nonstructural protein NS1), (NCVPI).
OS Bombus densonucleositis virus (BMDNV) (Silkworm densovirus).
OC Viruses; ssDNA viruses; Parvoviridae; Densovirinae; Iteravirus.
NCBI_TaxID=10809;
[1]
RN SEQUENCE FROM N.A.
RC STRAIN=Isolate INA;
RA MEDLINE=87112952; PubMed=3027382;
RX Barido H., Kusuda J., Gojobori T., Maruyama T., Kawase S.;
RT "Organization and nucleotide sequence of a densovirus genome imply a
RT host-dependent evolution of the parvoviruses.",
RL J. Virol. 61:553-560(1987).
CC -I- SIMILARITY: BELONGS TO THE PARVOVIRUSES NONCAPSID PROTEIN FAMILY.
CC
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (see http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC
CC EMBL; M15123; AAA67698.1; -.
CC DR PIR; G26796; VCEPVE2.
CC DR InterPro; IPR001257; Parvo.NS1.
CC DR Pfam; PF01057; Parvo.NS1; 1.
CC DR Nonstructural protein.
CC SO SEQUENCE 885 AA; 100958 MW; 98430617C534D255 CRC64;

Query Match 6.1%; Score 88.5; DB 1; Length 885;
Best Local Similarity 19.2%; Pred. No. 9.5;
Matches 61; Conservative 47; Mismatches 112; Indels 97; Gaps 13;

Oy 17 HIADFGSGRIPOFDALMPSEKLEQAMAEVPGLSAHPYISGYEKACFSMAVMEQ 76
Db 601 NVADII-KKALGTEKKNKTRPAPYPTTAMSEQFNISASSTGRDINNPI- 649
Oy 77 LDEGLPYAVEDEDVLLEGAEOFL-----AEDTWEERDKDSAFIVRLTFEAKVI 129
Db 650 VDSLPY---FENNPKRQVGIYDIYDINGTAYAKCKMEKRKPTNGLLIYASTLKGNY 706
Oy 130 ---VAPDKVLATENSPFLLESEHCAGTAYIISREAMPFTLDRAVLPPERIKA----- 180
Db 707 TPLAPFTNIMT---PIGLEN-----GYFMSNDQIERERBDLTTSVPDALTATKNS 756
Oy 181 -----VDLMMFTYFPDK--EGMPYQVSPALCTQELHAKTLLSONSMG----- 222
Db 757 ASNNINAAVVDYMGVYFEEQAKPQSPMKFMTGFEVIRND-----DSLNAKMDILI 808
Oy 223 -----SDLEKRDREGRRHRRSLKYMF-----DLKRALGKFGREKK-- 258
Db 809 KTRIRLTGLQSGREWVAFTDRIPQYFTSQYQFRYPNINIDPLLRISGTFGLPTKRPM 868
Oy 259 -----RMERQRAEL 268
Db 869 DSRIALGELQKORKMNL 885

```


THIS PAGE BLANK (USPTO)


```
QY 61 VERACPMASHAVIMEQALDDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFYR 120
DE LGTE.
GN LGTE.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=487;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=M978;
RX MEDLINE=21467954; PubMed=11583844;
RA Zhu P., Kitch M.J., Tsai C.-M.;
RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
Expression in Two L8-Immunotype Strains of Neisseria meningitidis.";
RL FEMS Microbiol. Lett. 203:173-177(2001).
DR EMBL: AF355193; AAL12842.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
SQ SEQUENCE 280 AA; 32898 MW; 504256FF5D92P97 CRC64;

RESULT 5
Q9EVD3 PRELIMINARY: PRT: 280 AA.
ID Q9EVD3
AC Q9EVD3:
DT 01-MAR-2001 (TReMBLrel. 16, Created)
DT 01-MAR-2001 (TReMBLrel. 16, Last sequence update)
DE Galactosyl transferase.
GN LGTE.
OS Neisseria subflava.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=28449;
RN (1)
RP SEQUENCE FROM N.A.
RX MEDLINE=21142520; PubMed=11208792;
RA Arking D., Tong Y., Stein D.C.;
RT "Analysis of lipooligosaccharide biosynthesis in the Neisseriaceae.";
RL EMBL: AF240672; AAG09766.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
DR Transferase.
SQ SEQUENCE 280 AA; 32762 MW; 85CF77E94F215842 CRC64;

Query Match 89.4%; Score 1287; DB 2; Length 280;
Best Local Similarity 90.4%; Pred. No. 2.2e-106;
Matches 253; Conservative 10; Mismatches 17; Indels 0; Gaps 0;

QY 1 MONHVISLASAERRAHIDTFEGSRGIPQFDPALMPSERLQMAELVPGLSAHPLYSG 60
DB 1 MONHVISLASAERRAHIDTFEGSRGIPQFDPALMPSERLQMAELVPGLSAHPLYSG 60
QY 61 VERACPMASHAVIMEQALDDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFYR 120
DB 61 VERACPMASHAVIMEQALDDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFYR 120
QY 121 LETMFAKVIIVRPDKVLYNENRSPFLLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
DB 121 LETMFAKVIIVRPDKVLYNENRSPFLLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
QY 121 LETMFAKVIIVRPDKVLYNENRSPFLLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
DB 121 LETMFAKVIIVRPDKVLYNENRSPFLLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
QY 181 VDMMFTYFEDKGMFVYOVSPALCTQELHYAKFLSNMGLSDEKDRQERRHRSIK 240
DB 181 VDMMFTYFEDKGMFVYOVSPALCTQELHYAKFLSNMGLSDEKDRQERRHRSIK 240
QY 241 VMEFDLKRALGKFGREKKRMEROROAELKRYGRRVILRK 280
DB 241 VMEFDLKRALGKFGREKKRMEROROAELKRYGRRVILRK 280

RESULT 6
Q93EK4 PRELIMINARY: PRT: 280 AA.
ID Q93EK4
AC Q93EK4:
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)

QY 61 VERACPMASHAVIMEQALDDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFYR 120
DE LGTE.
GN LGTE.
OS Neisseria meningitidis.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=487;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=M978;
RX MEDLINE=21467954; PubMed=11583844;
RA Zhu P., Kitch M.J., Tsai C.-M.;
RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
Expression in Two L8-Immunotype Strains of Neisseria meningitidis.";
RL FEMS Microbiol. Lett. 203:173-177(2001).
DR EMBL: AF355193; AAL12842.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
SQ SEQUENCE 280 AA; 32898 MW; 504256FF5D92P97 CRC64;

Query Match 89.0%; Score 1282; DB 2; Length 280;
Best Local Similarity 90.0%; Pred. No. 6e-106;
Matches 252; Conservative 9; Mismatches 19; Indels 0; Gaps 0;

QY 1 MONHVISLASAERRAHIDTFEGSRGIPQFDPALMPSERLQMAELVPGLSAHPLYSG 60
DB 1 MONHVISLASAERRAHIDTFEGSRGIPQFDPALMPSERLQMAELVPGLSAHPLYSG 60
QY 61 VERACPMASHAVIMEQALDDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFYR 120
DB 61 VERACPMASHAVIMEQALDDEGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFYR 120
QY 121 LETMFAKVIIVRPDKVLYNENRSPFLLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
DB 121 LETMFAKVIIVRPDKVLYNENRSPFLLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
QY 121 LETMFAKVIIVRPDKVLYNENRSPFLLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
DB 121 LETMFAKVIIVRPDKVLYNENRSPFLLESEHCGTAGYIISREARFPLDRPAVLPERRIKA 180
QY 181 VDMMFTYFEDKGMFVYOVSPALCTQELHYAKFLSNMGLSDEKDRQERRHRSIK 240
DB 181 VDMMFTYFEDKGMFVYOVSPALCTQELHYAKFLSNMGLSDEKDRQERRHRSIK 240
QY 241 VMEFDLKRALGKFGREKKRMEROROAELKRYGRRVILRK 280
DB 241 VMEFDLKRALGKFGREKKRMEROROAELKRYGRRVILRK 280

RESULT 7
Q93PR9 PRELIMINARY: PRT: 279 AA.
ID Q93PR9
AC Q93PR9:
DT 01-DEC-2001 (TReMBLrel. 19, Created)
DT 01-DEC-2001 (TReMBLrel. 19, Last sequence update)
DT 01-MAR-2002 (TReMBLrel. 20, Last annotation update)
DE Glycosyl transferase LgtB.
GN LGTB.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxId=485;
RN (1)
RP SEQUENCE FROM N.A.
RC STRAIN=PID2;
RA Tong Y., Arking D., Reinhold V., Stein D.C.;
RT "Characterization of lipooligosaccharide structures found in Neisseria
gonorrhoeae PID2.";
RL Submitted (OCT-2000) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF313394; AAK70339.1; -.
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
DR Transferase.
SQ SEQUENCE 279 AA; 31637 MW; 7297F13AFA732379 CRC64;

Query Match 72.7%; Score 1046.5; DB 2; Length 279;
Best Local Similarity 73.3%; Pred. No. 5.1e-85;
Matches 214; Conservative 18; Mismatches 35; Indels 25; Gaps 3;
```

QY	1	MONHVIASLSAAEERAHIAHDPFGSKGIPDFOEFDLMSERLEQMAALVPELSAHPYLSG	60
Db	1	MONHVIASLSAAERRAHIAATFGSRGIPDFOFDLMSERLEQMAALVPELSAHPYLSG	60
QY	61	VEKACEMSHAVLMEOALDEGLPYIAVEDDVLGEGAEQFLAEDPTWLTEERDKDSAEIVR	120
Db	61	VEKACEMSHAVLMEOALDEGVPIYIAVEDDVLGEGAEQFLAEDPTWLTEERDDPSAEIFVR	120
QY	121	LETMPAKYIVAPDKVVLNENSPFLBSEHCGTAGYIISRAMPFLDPRFVILPERIKA	180
Db	121	LETMPMBHVLITPSGADYGGRAFLPFLBSEHCGTAGYIISRAMPFLDPRFVILPAERIKIA	180
QY	181	VDLMIAFTYFDPKEGMPYQVSPALCTOELAHAKFLSONSMSEGLDEKDRQGR-----	233
Db	181	VDLMFMGSPDDREGMPYQQLPALCAQELAHYAKFHDQNSALGSLLEHDRRLNRKQWRDS	240
QY	234	-----RHRRSLKVMFDLKRALGKGRGRRKKMMEQRQAELEKUYGRVILFK	280
Db	241	PANTEFKH-----LIRALTKIGREKREKRRK-----CQÖTKGILIVPQÖ	279

RESULT 8			
093JE5			
ID	093JE5	PRELIMINARY;	PRT; 275 AA.
AC	Q93JE5		
DT	01-DEC-2001 (TrEMBLrel. 19, Created)		
DT	01-DEC-2001 (TrEMBLrel. 19, Last sequence update)		
DT	01-MAR-2002 (TrEMBLrel. 20, Last annotation update)		
DE	LetB.		
GN	LetB.		
OS	Neisseria meningitidis.		
OC	Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.		
NC	NCBI_TaxID=487;		
RN	[1]		
RP	SEQUENCE FROM N.A.		
RC	STRAT=MG78;		
RC	MEDLINE=21467954; PubMed=11583844;		
RA	Zhu P., Klutch M.J., Tsai C.-M.;		
RT	"Genetic Analysis of Conservation and Variation of Lipooligosaccharide		
RL	Expression in Two L8-Immunotype Strains of Neisseria meningitidis.";		
DR	EMBL: AF355193; AAL12841.1; -.		
DR	InterPro: IPR002654; GT_25.		
DR	Pfam: PF01755; Glyco_Transf_25; 1.		
Q0	SEQUENCE 275 AA; A73037E0625EE7B3 CRC64;		

Query Match	70.48;	Score 1014;	DE 2;	Length 275;
Best Local Similarity	75.58;	Pred. No. 3.9e-82;		
Matches 203; Conservative	21;	Mismatches 41;	Indels 4;	Gaps 1

QY	1	MÖNNVHSLASAAERRAHIAADFEGRGIPPOEFQDLMSPSELEOMATVYGLSAAHPTLG	60
QY	1	MÖNNVHSLASAAERRAHIAATFGANGIPPOFDLMSPSERIEOMAAVLVGLSHNPPLSG	60
QY	61	VEKACEMSHAVLMQALDEGLIPYIAVEDDVLVLEGAEQFLAEDVTWLEERFDDSDAIVR	120
QY	61	VEKACEMSHAVLMQALDEGLIPYIAVEDDVLVLEGAEQFLAEDVTWLEERFDDSDAIVR	120
QY	61	VEKACEMSHAVLMQALDEGLIPYIAVEDDVLVLEGAEQFLAEDVTWLEERFDDSDAIVR	120
QY	121	LETAPAKVIYAPDKVLVLEANSFPLTSEHCGTAGYIIISEAMFPLIDRAVILPPEKIRA	180
QY	121	LETAPAKVIYAPDKVLVLEANSFPLTSEHCGTAGYIIISEAMFPLIDRAVILPPEKIRA	180
QY	121	LETAPAKVIYAPDKVLVLEANSFPLTSEHCGTAGYIIISEAMFPLIDRAVILPPEKIRA	180
QY	181	VDLMEATFFEDKEGMPYQVSPALCTQELHNAKPLSONSMGLSDLEKD---REOGRHR	236
QY	181	VDLMEATFFEDKEGMPYQVSPALCTQELHNAKPLSONSMGLSDLEKD---REOGRHR	236
QY	181	VDLMEATFFEDKEGMPYQVSPALCTQELHNAKPLSONSMGLSDLEKD---REOGRHR	236
QY	237	RSIAVMEDLKALCKFGREKKRRERORQ	265
QY	241	PANTFKRRILRALTKISREKREKRRORRQ	269

RESULT 9
Q9JW65

ID O9JW65; PRELIMINARY; PRT; 268 AA.
 DC O9JW65;
 DT 01-OCT-2000 (TrEMBLrel. 15, Created)
 DT 01-OCT-2000 (TrEMBLrel. 15, last sequence update)
 DT 01-MAR-2002 (TrEMBLrel. 20, last annotation update)
 DE Lacto-N-neotetraose biosynthesis glycosyl transferase (LgtH).
 GN LgtB2 OR NMA0527 OR LgtH.
 OS Neisseria meningitidis (serogroup A), and
 OS Neisseria meningitidis.
 OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
 OX NCBI_Taxid=56599, 487;
 [1]
 RP SEQUENCE FROM N.A.
 RC SPECIES=N.meningitidis (serogroup A);
 RC STRAIN=Z2491 / SEROGROUP A; SEROTYPE 4A;
 RX MEDLINE=20222556; Pubmed=10761919;
 RA Parkhill J., Achtman M., James K.D., Bentley S.D., Churcher C.,
 RA Klee S.R., Morelli G., Basham D., Brown D., Chillingworth T.,
 RA Davies R.M., Davis P., Dealla K., Fellwell T., Hamlin N., Holtroyd S.,
 RA Jagers K., Leather S., Moule S., Mungall K., Quail M.A.,
 RA Rajandream M.A., Rutherford K.M., Simmonds M., Skelton J.,
 RA Whitehead S., Spratt B.G., Barrall B.G.,
 RT "Complete DNA sequence of a serogroup A strain of Neisseria
 RT meningitidis Z2491.";
 RL Nature 404:502-506(2000).

PC SPECIES=N.meningitidis; STRAIN=A1;
RX MEDLINE=21467954; PubMed=11583844;
RA Zhu P., Kitchin M.D., Tsai C.-M.;
RT "Genetic Analysis of Conservation and Variation of Lipooligosaccharide
RT Expression in Two L8-Immunotype Strains of Neisseria meningitidis";
RL FEMS Microbiol. Lett. 203:173-177(2001).
DR EMBL: AL162753; CAB83819.1; -
DR EMBL: AF355194; AL12844.1; -
DR InterPro: IPR002654; GT_25.1;
DR Pfam: PF01755; Glyco_Transf_25; 1.
KW Complete proteome.
SQ SEQUENCE 268 AA; 30444 MW; 8FA3967D053DD02 CRC64;

Query Match	69.9%;	Score 1007;	DB 16;	Length 268;
Best Local Similarity	72.9%;	Pred. No. 1.6e-81;		
Matches 204; Conservative	24;	Mismatches 40;	Indels 12;	Gaps 2

Qy	1	MONVITSLASAAERRAHNTADPFSSRGJPEOPFDIMAMSERLEQAMAEVLGSLAHPLSG	60
Dp	1	MONVITSLASAAERRAHNTDFFVGKGPJPFQFDALMBSERLEQAMAEVLGSLAHPLSG	60
Qy	61	VEKACFMSHAVALMBOALDEBGLPYIAVEDDVLIGEGBQFLAEDBTWLEERFDKXSAFIVR	120
Dp	61	VEKACFMSHAVALMBOALDEBGLPYIAVEDDVLIGEGBQFLAEDBTWLEERFDKXSAFIVR	120
Qy	121	LETMFPAKIYAPDKVLYLNTENNSFFLSEHCGTGYIISIRPAMFFLDRPAVLPRPEIKTA	180
Dp	121	LETMFPAKIYAPDKVLYLNTENNSFFLSEHCGTGYIISIRPAMFFLDRPAVLPRPEIKTA	180
Qy	181	VDLMNFTFFPEKEGMPYQVSPALCTOELHNAKFLSONSMGLSDEKDRQGRHRRSLK	240
Dp	181	VDLMNFTFFPEKEGMPYQVSPALCTOELHNAKFLSONSMGLSDEKDRQGRHRRSLK	240
Qy	241	VMFIDKRALGFGREKKRMERORQALEKYVGRVILFK	280
Dp	241	VMFIDKRALGFGREKKRMERORQALEKYVGRVILFK	280
Qy	291	VL--LKRALGIGREIEERAREKROKTEKLTGHNHVPFE	268
Dp	291	VL--LKRALGIGREIEERAREKROKTEKLTGHNHVPFE	268

RESULT 10	
Q9RCN1	
ID Q9RCN1	PRELIMINARY; PRT; 248 AA.
AC Q9RCN1;	
DT 01-MAY-2000 (TREMBLrel. 13, Created)	
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)	
DT 01-MAR-2002 (TREMBLrel. 20, Last annotation update)	
DE Glycosyltransferase.	

GN LGTB.
OS Neisseria gonorrhoeae.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=485;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=1291;
RX MEDLINE=20305049; PubMed=10844691;
RA Harvey H.A., Porat N., Campbell C.A., Jennings M., Gibson B.W.,
Phillips N.J., Apicella M.A., Balke M.S.;
RT "Gonococcal lipooligosaccharide is a ligand for the asialoglycoprotein
receptor on human sperm";
RL Mol. Microbiol. 36:1059-1070(2000).
DR EMBL: AF121135; AAF14360.1; -
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
KW Transferase.
SQ SEQUENCE 248 AA; 28036 MW; D9036FE07D8C9397 CRC64;

Query Match 69.5%; Score 1001; DB 2; Length 248;
Best Local Similarity 82.2%; Pred. No. 4.9e-81;
Matches 194; Conservative 15; Mismatches 27; Indels 0; Gaps 0;
QY 1 MONHVISLASAERRAHIDTSGSGIPQFDPALMPSERLDQAMAEIVPGLSAHPYLSG 60
DB 1 MONHVISLASAERRAHIDTSGSGIPQFDPALMPSERLDQAMAEIVPGLSAHPYLSG 60
QY 61 VEKACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120
DB 61 VEKACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120
QY 121 LETMFAKVIIRPDKVLNENRSEFPLESEHCCTAGYIISREAMRFLDRFAVLPERRIKA 180
DB 121 LETMFAKVIIRPDKVLNENRSEFPLESEHCCTAGYIISREAMRFLDRFAVLPERRIKA 180
QY 181 VOLAMFTYFEDEKGMVYOVSPALCTQELHYAKFLSQNSMLGSDLEKDEOGRHR 236
DB 181 VOLAMFTYFEDEKGMVYOVSPALCTQELHYAKFLSQNSMLGSDLEKDEOGRHR 236
QY 181 VOLAMFTYFEDEKGMVYOVSPALCTQELHYAKFLSQNSMLGSDLEKDEOGRHR 236
DB 181 VOLAMFTYFEDEKGMVYOVSPALCTQELHYAKFLSQNSMLGSDLEKDEOGRHR 236

RESULT 11
Q9EVD4 PRELIMINARY; PRT; 275 AA.
ID Q9EVD4
AC Q9EVD4;
DT 01-MAR-2001 (Tremblrel. 16, Created)
DT 01-MAR-2001 (Tremblrel. 16, last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
DE Galactosyl transferase.
GN LGTB.
OS Neisseria subflava.
OC Bacteria; Proteobacteria; beta subdivision; Neisseriaceae; Neisseria.
OX NCBI_TaxID=28449;
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=21142520; PubMed=11208792;
RA Arking D., Tong Y., Stein D.C.;
RT "Analysis of lipooligosaccharide biosynthesis in the Neisseriaceae";
RL J. Bacteriol. 183:934-941(2001).
DR EMBL: AF240672; AAG09765.1; -
DR InterPro: IPR002654; GT_25.
DR Pfam: PF01755; Glyco_transf_25; 1.
KW Transferase.
SQ SEQUENCE 275 AA; 31441 MW; D719F3815F64D14C CRC64;

Query Match 68.3%; Score 984; DB 2; Length 275;
Best Local Similarity 73.6%; Pred. No. 1.8e-79;
Matches 198; Conservative 21; Mismatches 46; Indels 4; Gaps 2;
QY 1 MONHVISLASAERRAHIDTSGSGIPQFDPALMPSERLDQAMAEIVPGLSAHPYLSG 60
DB 1 MONHVISLASAERRAHIDTSGSGIPQFDPALMPSERLDQAMAEIVPGLSAHPYLSG 60
QY 61 VEKACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120

DB 61 VEKACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVR 120
QY 121 LETMFAKVIIRPDKVLNENRSEFPLESEHCCTAGYIISREAMRFLDRFAVLPERRIKA 180
DB 121 LETMFAKVIIRPDKVLNENRSEFPLESEHCCTAGYIISREAMRFLDRFAVLPERRIKA 180
QY 181 VOLAMFTYFEDEKGMVYOVSPALCTQELHYAKFLSQNSMLGSDLEKDEOGRHR--RRS 238
DB 181 VOLAMFTYFEDEKGMVYOVSPALCTQELHYAKFLSQNSMLGSDLEKDEOGRHR--RRS 238
QY 239 LKYMED--LKRALGKFGREKKRMRERQ 265
DB 241 PANTFERRLIRALTISRERERKRORREQ 269

RESULT 12
Q8RKU0 PRELIMINARY; PRT; 266 AA.
ID Q8RKU0
AC Q8RKU0;
DT 01-JUN-2002 (Tremblrel. 21, Created)
DT 01-JUN-2002 (Tremblrel. 21, last sequence update)
DT 01-JUN-2002 (Tremblrel. 21, last annotation update)
DE Putative galactosyl transferase Lic2B.
OS Haemophilus influenzae.
OC Bacteria; Proteobacteria; gamma subdivision; Pasteurellaceae;
OX NCBI_TaxID=727;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=6622;
RA Pettigrew M.M., Foxman B., Marrs C.F., Gilsdorf J.R.;
RT "Identification of LOS biosynthesis gene lic2B as a putative virulence
factor in strains of non-typable Haemophilus influenzae that cause
otitis medi.";
RT Submitted (MAR-2002) to the EMBL/Genbank/DBJ databases.
DR EMBL: AY091470; AAM12037.1; -
KW Transferase.
SQ SEQUENCE 266 AA; 31182 MW; B21CFA372200E2DC CRC64;

Query Match 35.5%; Score 510.5; DB 2; Length 266;
Best Local Similarity 41.4%; Pred. No. 2.1e-37;
Matches 116; Conservative 43; Mismatches 96; Indels 25; Gaps 7;
QY 3 NHVISLASAERRAHIDTSGSGIPQFDPALMPSERLDQAMAEIVPGLSAHPYLSGVE 62
DB 4 NHVISLASAERRAHIDTSGSGIPQFDPALMPSERLDQAMAEIVPGLSAHPYLSGVE 62
QY 63 KACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVRLE 122
DB 63 KACFMSHAVIMEQALDGLPYIAVEEDVLLGEGAEQFLAEDTWLEERFDKDSAFIVRLE 122
QY 123 TFMFAKVIIRPDKVLNENRSEFPLESEHCCTAGYIISREAMRFLDRFAVLPERRIKA 182
DB 123 TFMFAKVIIRPDKVLNENRSEFPLESEHCCTAGYIISREAMRFLDRFAVLPERRIKA 182
QY 183 LKMFYTFEFGKGMVYOVSPALCTQELHYAKFLSQNSMLGSDLEKDEOGRHR--GRRHR 240
DB 183 LKMFYTFEFGKGMVYOVSPALCTQELHYAKFLSQNSMLGSDLEKDEOGRHR--GRRHR 240
QY 241 VMEFLKRALGKFGREKKRMRERORAELEKVGRRYILRK 280
DB 241 VMEFLKRALGKFGREKKRMRERORAELEKVGRRYILRK 280
QY 236 TVL-----ISLAKGPKK-----ILKRIY-RKLFISK 260
DB 236 TVL-----ISLAKGPKK-----ILKRIY-RKLFISK 260

RESULT 13
Q9L7A4 PRELIMINARY; PRT; 280 AA.
ID Q9L7A4
AC Q9L7A4;
DT 01-OCT-2000 (Tremblrel. 15, Created)
DT 01-OCT-2000 (Tremblrel. 15, last sequence update)
DT 01-MAR-2002 (Tremblrel. 20, last annotation update)
DE Galactosyltransferase II.

THIS PAGE BLANK (USPTO)